

T.C
DOKUZ EYLÜL ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
EKONOMETRİ ANABİLİM DALI
DOKTORA TEZİ

**ZAMAN SERİLERİNDE
SAPAN DEĞERLERİN ANALİZİ
ÜZERİNE BİR ARAŞTIRMA**

Hazırlayan

Ahmet KAYA

Danışman

Doç.Dr. Levent ŞENYAY

**İzmir
1999**

**T.C. YÜKSEKÖĞRETİM KURULU
DOKÜMANTASYON MERKEZİ**

T.C
SOSYAL BİLİMLER ENSTİTÜSÜ
EKONOMETRİ ANABİLİM DALI
DOKTORA TEZİ

**ZAMAN SERİLERİNDE
SAPAN DEĞERLERİN ANALİZİ
ÜZERİNE BİR ARAŞTIRMA**

89369

Hazırlayan

Ahmet KAYA

Danışman

Doç.Dr. Levent ŞENYAY

**İzmir
1999**

YEMİN

Doktora Çalışması olarak sunduğum "Zaman Serilerinde Sapan Değerlerin Analizi Üzerine Bir Araştırma" adlı çalışmanın, tarafımdan, bilimsel ahlak ve geleneklere aykırı düşecek bir yardıma başvurmaksızın yazıldığını ve yararlandığım eserlerin bibliyografyada gösterilenlerden oluştuğunu, bunlara atıf yapılarak yararlanılmış olduğunu belirtir ve bunu onurumla doğrularım.

Ahmet Kaya
Ahmet KAYA

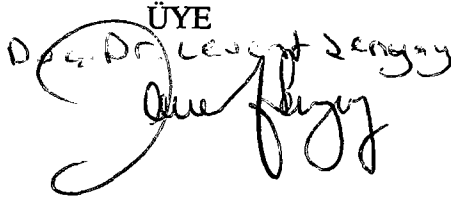
TUTANAK

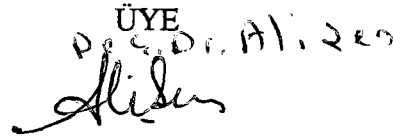
Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsünün 13. /01/1999 tarih ve 26 sayılı toplantısında oluşturulan jüri, Lisansüstü Öğretim Yönetmeliğinin 3. maddesine göre Ehenen'de Anabilim Dalı Doktora öğrencisi Ahmet Kaya'nın Roman Serilerinde Sosyal Değerler Analizi konulu tezi incelenmiş ve aday 13./01/1999 tarihinde Saat 10.00'da jüri önünde tez savunmasına alınmıştır.

Adayın kişisel çalışmaya dayanan tezini savunmasından sonra 90... dakikalık süre içinde gerek tez konusu, gerekse tezin dayanağı olan anabilim dallarından jüri üyelerince sorulan sorulara verdiği cevaplar değerlendirilerek tezin başarılı olduğuna oy birliği ile karar verildi.

Prof. Dr. Fikret İkin
BAŞKAN



ÜYE
Doç. Dr. Levent Zengin


ÜYE
Doç. Dr. Ali Zengin


YÜKSEKÖĞRETİM KURULU DÖKÜMANTASYON MERKEZİ
TEZ VERİ FORMU

Tez No :

Konu Kodu :

Üniv. Kodu

Tez Yazarının

Soyadı : KAYA.

Adı : Ahmet.

**Tezin Türkçe Adı : ZAMAN SERİLERİNDE SAPAN DEĞERLERİN
ANALİZİ ÜZERİNE BİR ARAŞTIRMA.**

Tezin Yabancı Dildeki Adı : A RESEARCH ON OUTLIER IN TIME SERIES.

Tezin Yapıldığı

Üniversite : Dokuz Eylül.

Enstitü : Sosyal Bilimler.

Yıl : 1999.

Tezin Türü : Doktora.

Dili : Türkçe.

Sayfa Sayısı : 100.

Referans Sayısı : 50.

Tez Danışmanınının

Ünvanı Adı ve Soyadı : Doç.Dr. Levent ŞENYAY.

Türkçe Anahtar Kelimeler :

- 1-Zaman Serileri.
- 2-ARIMA.
- 3-Sapan Değerler.
- 4-Zaman Serilerinde Sapan Değerler.
- 5-Varyans Analizi

İngilizce Anahtar Kelimeler :

- 1-Time Series.
- 2-ARIMA.
- 3-Outliers.
- 4-Outliers in Time Series.
- 5-Analysis of Variance.

Tarih :

İmza : Ahmet Kaya.

ÖZET

Çalışmada incelenen sapan değerler (Outliers) kavramı, örneklemdaki diğer gözlemlerden farklı olan ve tahmin sonuçlarını büyük ölçüde etkileyen bir ya da birden çok gözlem değeridir.

Sapan değerler; model yanlışlıkları, gerekli bazı dönüşümlerin yapılmamış olması, ölçüm, tartım ve kaydetme hataları gibi çeşitli nedenlerden kaynaklanabilmektedir. Örneklemdaki bazı gözlemlerin genelden farklı olmalarının çeşitli sebepleri olabilir. Sapan değerler doğal rasgelelik sonucunda ortaya çıkabildikleri gibi kişilerin, makinaların hatalarından veya benzer nedenlerden oluşabilmektedir.

Sapan değerler ekonomik anlamda; grevler, savaşlar, bir malın piyasasında veya fiziksel sistemde meydana gelebilecek beklenmeyen değişiklikler ve benzer durumlardan oluşabilir. Belirtilen durumlar beklenmeyen müdahale olarak tanımlanmaktadır.

Bu çalışmada, Chang Tiao ve Chen (1988) tarafından yapılan simülasyon çalışmaları kullanılarak elde edilmiş tablolar, yeniden planlanmış ve planlama sonucunda modelleme işlemi gerçekleştirilmiştir. Yapılan varyans analizi işlemi sonucunda sapan değer tespitinde; seri uzunluğu, seri yapısı, sapan değer büyüklüğü, sapan değer tipleri ve model farklılıklarının sapan değer tespitinde ne derece etkili olduğu ortaya konmuştur. Ayrıca tarama süreçlerinin etkinliklerini ortaya koymak açısından benzer çalışmalar yapılmış ve gerekli çıkarımlar sunulmuştur.

Yapılan ($2^3 3^2$) özel seçimli, çok etkenli deney düzeninden oluşan varyans analizi işlemi sonucunda; sapan değer tespitinde model farklılığının (AR, MA), sapan değer büyüklüğünün (3σ , 5σ), seri genişliğinin (50, 100, 150), sapan değer türünün (AO(Additive Outlier), IO(Innovational Outlier)) önemli, ölçüt değer duyarlılığının ($C=3.00$, 3.50, 4.00) önemsiz, olduğu sonucuna varılmıştır.

Dikkati çeken diđer bir husus, seri ierisinde sapan deđer sayısının artışı tarama srelerinin glerinde sapan deđer tespit ynnden nemli oranda g kaybı dođurmuştur.

Ayrıca, zel seimli ok etkenli deney dzeninden oluřan varyans analizi iřlemi ile sapan deđer tarama srelerinin, sapan deđer tespit gc ynnden farklı oldukları ortaya konmuştur. Sz konusu sreler ierisinde "Ardışık Yntem" (Iterative Procedure) olarak bilinen yntem en etkili olduđundan teorik yapısı ile tanımlanmış ve bilgisayar programına dnřtrlmřtr.



SUMMARY

The concept of outliers investigated in this study have been regarded as the observations which are extremely different from the others in a sample or one or two observations which effect the estimation and forecasting process in a great extent.

The outliers may result from several reasons such as; lack of necessary transformations, the errors made on observations and not being able to assume the correct model. In other words, some observations may be very “different” for different reasons other than the naturel randomness.

The outliers in economic investigations may happen after expected or unexpected consequences.

The data used in this study were obtained from the simulation experiments performed on several time series model by *Chang, Tiao and Chen* (1988). The original simulation results were reorganized and remodeled under a suitable experimental design. Two different time series models (AR, MA) with two outliers types (AO, IO) in two different size (3σ , 5σ) were arranged in three different time series size (50, 100, 150) and different sensitivity coefficient (3.00, 3.50, 4.00). The type of factorial design was ($2^3 3^2$) that is 3 factors each at 2 levels and two factors at 3 levels each.

The result of the analysis of variance performed for this design indicated that the main effect except that of sensitivity coeffecient were statistically significant. On the other hand the iterative outlier search procedure was found to be most effective search process. A computer program was developed for this effective search procedure.

ÖNSÖZ

Belirsizlik durumlarında karar vermek ve geleceğe dönük projeksiyonlar yapmak için vazgeçilmez analiz aracı haline gelen istatistik yöntemlerde değişim ve gelişimler çok hızlı değildir. Bu durum, kullanılan analiz araçlarının etkinliklerini artırmayı gerekli kılmaktadır. Bunun için veri setleri üzerinde analiz çalışmasına başlamadan önce dönüşüm ve düzeltme işlemleri gerekebilir. Verilerin özellikleri gereği uygulanması gereken bir çok dönüşüm ve düzeltme yöntemi mevcuttur. Bu tip işlemler genelde sayısal olarak birbirine yakın homojen veriler setleri üzerinde yapılmaktadır. Eğer veriler gözlem değerlerinin veya hata değerlerinin doğrusal bir kombinasyonu olarak elde edilen zaman serisi modellerine uyduruluyorsa, diğer gözlem değerlerinden farklı olan bir veya birden çok gözlemin parametre değerlerine etkileri büyük olmaktadır.

Parametre değerlerinin gerçeğe yakın ve yansız olmaları, seride bulunan bu tip verilere ilişkin hata terimleri üzerinde düzeltme işlemini zorunlu kılmaktadır. Tanımı yapılan bu analiz işlemine sapan değer analizi denir.

İstatistik analiz yöntemleri olarak kullanılan regresyon analizi, çok değişkenli istatistik analiz, deneme planlama ve varyans analizi ve diğer tahmin yöntemleri için; araştırma yayını, makale, yüksek lisans ve doktora düzeyinde sapan değer analizi üzerine çok sayıda çalışma mevcuttur. Teorik zorluklar nedeniyle çok fazla çalışılmayan zaman serilerinde sapan değerler kavramı 1972 yılında İngiliz *Fox* ile başlamış, yurt dışında özellikle Amerikan, Japon, Çin, Kore ve Hintli bilim adamlarının ilgi odağı haline gelmiş, diğer ülke araştırmacıları tarafından da artan oranda çalışılmaya başlanmıştır. Bu konuda *Chang*, Wisconsin Üniversitesi'nde doktora çalışması yapmıştır. Ülkemizde özellikle kalite kontrol amaçlı süreçler için makale düzeyinde çalışılan konu, tarafımdan Ege üniversitesi Fen Bilimleri Enstitüsü İstatistik Teorisi Anabilim Dalı'nda kalite kontrol amaçlı süreçler bazında yüksek lisans düzeyinde çalışılmıştır.

Doktora tez konusu olarak belirlenen bu çalışmada, sapan değer tespitinde etken oldukları düşünölen zaman serileri modelleri, sapan değer büyüklükleri, seri uzunlukları, ölçüt değer duyarlılıkları, sapan değer türleri, sapan değer sayıları ve tarama süreçlerinin performansları gibi faktörler üzerinde *Chang, Tiao ve Chen (1988)* tarafından yapılan simölasyon tabloları kullanılarak varyans analizi yöntemi ile teorik çıkarsama ve bulgular elde edilmiştir.

Uygulama bazında; kalite kontrol amaçlı, veri setini düzeltme amaçlı ve parametre tahminlerinde etkinlik artırma amaçlı olmak üzere üç farklı veri seti üzerinde analiz işlemleri yapılmış, çıkan sonuçlar teorik bulgularla birlikte ifade edilmiştir.

Çalışmalarımı sonuçlandırmama katkı yapan hocalarımdan; doktora öğrencisi olduğum süre içerisinde danışmanlık görevini üstlenen ve çalışmalarımıda yardımcı olan Doç.Dr. Levent ŞENYAY'a, değerli katkıları ile beni yönlendiren Doç.Dr Ali ŞEN'e, bütün çalışmalarımıda olduğu gibi doktora çalışmama da katkı yapan Prof.Dr. Fikret İKİZ'e, her zaman yanımda gördüğüm çalışmalarımın varyans analizi kısmına destek veren değerli hocam, Doç.Dr. Serdar KURT'a, tez konusu belirleme aşamalarında görüşleri ile destek veren Hacettepe Üniversitesinden Prof. Dr. Cenap ERDEMİR'e, Doç. Dr. Mustafa GÜNEŞ'e, Dokuz Eylül Üniversitesi İktisadi ve İdari Bilimler Fakültesi Ekonometri Bölümü öğretim üyelerine ve yardımcılarına, katkı yapan herkese teşekkürlerimi ve saygılarımı sunarım.

Çalışmalarıma manevi desteklerini eksik etmeyen doktora çalışmalarım için gerekli zamanı yaratmama katkı yapan, mensubu bulunduğum Ege Üniversitesi Tire Meslek Yüksekokulu'nun değerli müdür, öğretim üye ve görevlilerine teşekkürlerimi sunarım.

Ayrıca zorlu çalışmalarımıda daima yanımda gördüğüm manevi yardım ve desteklerini eksik etmeyen sevgili eşim Neşe'ye teşekkür ederim.

İÇİNDEKİLER

BİRİNCİ BÖLÜM

	Sayfa
1. GİRİŞ	1
1.1. Kaynak Taramaları	4
1.2. İlişkili Çalışmalar	8

İKİNCİ BÖLÜM

2. ZAMAN SERİLERİ	10
2.1. Box-Jenkins Tahmin Yöntemi	16
2.1.1 AR Modeller	17
2.2.1 MA Modeller	20
2.3.1 ARMA Modeller	22
2.4.1 ARIMA Modeller	23
2.2. Modelin Uygunluk Testi	25
2.3. Box-Jenkins Yönteminin Üstün ve Zayıf Yönleri	26
2.3.1. Üstün Yönleri	26
2.3.2. Zayıf Yönleri	26
2.4. Zaman Serilerinin Analizi	27

ÜÇÜNCÜ BÖLÜM

3. SAPAN DEĞERLER	29
3.1. Temel Tanımlar	29
3.2. Kurlsız İlk Yaklaşımlar	31
3.3. Basit Örneklemeler için Sapan Değerler	38
3.4. Zaman Serilerinde Sapan Değerler	41
3.4.1. Müdahale Analiz Yöntemi	43
3.4.2. Ardışık Yöntem	52
3.4.3. Sapan Değer Tespit Ölçütleri ve Olasılığı	54
3.4.4. Tarama Yöntemleri için Performans Analizi	58

DÖRDÜNCÜ BÖLÜM

4. SAPAN DEĞER TESPİTİNDE FAKTÖR ANA ETKİLERİNİN ÖNEM DERECELERİNİN VARYANS ANALİZİ İLE BELİRLENMESİ VE UYGULAMA	62
4.1. Tek Sapan Değerlere İlişkin Varyans Analizi Çalışması	62
4.2. Çift Sapan Değerlere İlişkin Varyans Analizi Çalışması	71
4.3. Tarama Süreçlerine İlişkin Varyans Analizi Çalışması	77
4.4. Uygulama	84
4.4.1. Kontrol Amaçlı	85
4.4.2. Düzeltme Amaçlı (Temel Analiz Gereği)	92
4.4.3. Etkinlik Artırma Amaçlı	94

BEŞİNCİ BÖLÜM

5. SONUÇLAR	96
BİBLOGRAFYA VE EKLER	100
BİBLOGRAFYA	101
EK-1. ARDIŞIK YÖNTEMİN BİLGİSAYAR PROGRAMI	106
EK-2. DAĞILIMLARI VERİLEN ÖRNEKLER İÇİN SAPAN DEĞER İSTATİSTİKLERİ	109

TABLolar VE GRAFİKLER

ÜÇÜNCÜ BÖLÜM

Sayfa

Tablo-3.1, C=4.0, 1000 Tekrarlı, Sapan Değer Tarama Olasılıkları	[56]
Tablo-3.2, C=3.5, 1000 Tekrarlı, Sapan Değer Tarama Olasılıkları	[56]
Tablo-3.3, C=3.0, 1000 Tekrarlı, Sapan Değer Tarama Olasılıkları	[57]
Tablo-3.4, Tek Sapan Değer (IO) İçin Süreçlerin Performansı	[58]
Tablo-3.5, Tek Sapan Değer (AO) İçin Süreçlerin Performansı	[58]
Tablo-3.6, Çift Sapan Değer (AO) için Süreçlerin Performansı	[59]
Tablo-3.7, Çift Sapan Değer (IO) için Süreçlerin Performansı	[59]
Tablo-3.8, Üç Sapan Değer (AO) için Süreçlerin Performansı	[60]
Tablo-3.9, Üç Sapan Değer (IO) için Süreçlerin Performansı	[60]
Tablo-3.10, Üç Sapan Değer (Karışık) için Süreçlerin Performansı	[61]

DÖRDÜNCÜ BÖLÜM

Tablo-4.1, Tek Sapan Değer Tarama Olasılıkları	[63]
Tablo-4.2, (4.1) Modeline İlişkin Tasarım Çizelgesi	[64]
Tablo-4.3, (4.1) Modeline İlişkin Gözlem, Tahmin ve Hata Değerleri	[66]
Tablo-4.4, (4.1) Modeline İlişkin Varyans Analizi Sonuçları	[67]
Tablo-4.5, Tek Sapan Değer Verileri için Ana Etkilere İlişkin VA Sonuçları	[68]
Tablo-4.6, Tek Sapan Değer Verileri için Büyüklük-Model Etkileşimi	[69]
Tablo-4.7, Tek Sapan Değer Verileri için Tür-Model Etkileşimi	[69]
Tablo-4.8, Tek Sapan Değer Verileri için Büyüklük-Tür Etkileşimi	[70]
Tablo-4.9, Tek Sapan Değer Verileri için Tür-Genişlik Etkileşimi	[70]
Tablo-4.10, Tek Sapan Değer Verileri için Tür-Duyarlılık Etkileşimi	[70]
Tablo-4.11, AR Modelde İki Sapan Değer Tarama Olasılıkları	[71]
Tablo-4.12, (4.2) Modeline İlişkin Tasarım Çizelgesi	[72]
Tablo-4.13, (4.2) Modeline İlişkin Gözlem, Tahmin ve Hata Değerleri	[74]
Tablo-4.14, (4.2) Modeline İlişkin İki Sapan Değer Tarama Verileri için VA ...	[74]
Tablo-4.15, AR Modelde İki Sapan Değer Tarama Verileri için Ana Etkilere ...	[75]

Tablo-4.16 , Tarama Süreçlerine Bağlı Elde Edilen İlişkin Hata Kareler Ort._____	[78]
Tablo-4.17 , (4.3) Modeline İlişkin Tasarım Çizelgesi_____	[78]
Tablo-4.18 , (4.3) Modeline İlişkin Gözlem, Tahmin ve Hata Değerleri_____	[80]
Tablo-4.19 , Tarama Süreçlerine Bağlı Elde Edilen HKO verileri için VA Son.____	[80]
Tablo-4.20 , Tarama Süreçleri ve Diğer Faktör Ana Etkilerine İlişkin VA Son.____	[81]
Tablo-4.21 , Hata Kareler Ortalaması Verileri için Tür-Adet Etkileşimi_____	[82]
Tablo-4.22 Hata Kareler Ortalaması Verileri için Tür-Süreç Etkileşimi_____	[82]
Tablo-4.23 , Hata Kareler Ortalaması Verileri için Adet-Süreç Etkileşimi_____	[82]
Tablo-4.24 , Kontrol Amaçlı Box-Jenkins A Serisi Gözlem Değerleri (197 Adet) _	[85]
Tablo-4.25 , Box-Jenkins A Serisi için Değişik İterasyonlardaki Parametre Tah.____	[86]
Tablo-4.26 , Box-Jenkins A Serisi için Parametre Tahmin Değerleri_____	[86]
Tablo-4.27 , Box-Jenkins A Serisi Gözlem, Tahmin ve Hata Değerleri_____	[87]
Tablo-4.28 , Box-Jenkins A Serisi Gözlem, Tahmin ve Hata Değerleri(Devam)____	[88]
Tablo-4.29 , Box-Jenkins A Serisi Gözlem, Tahmin ve Hata Değerleri(Devam)____	[89]
Tablo-4.30 , Box-Jenkins A Serisi Kuşku Gözlem Serisi (59-69. Gözlemler)____	[89]
Tablo-4.31 , Box-Jenkins A Serisi Kuşku Gözlem Serisi (38-48. Gözlemler)____	[90]
Tablo-4.32 , Box-Jenkins A Serisi için Sapan Değer Analizi Sonuçları_____	[90]
Tablo-4.33 , Box-Jenkins A Serisi için Sapan Değer Analizi Sonuçları_____	[91]
Tablo-4.34 , İzmir ÇİMENTAŞ T.A.Ş'nin 1980-1987 Yılı İhracat Rakamları_____	[92]
Tablo-4.35 , İzmir ÇİMENTAŞ T.A.Ş Verileri için Değişik İterasyon Par..._____	[92]
Tablo-4.36 , İzmir ÇİMENTAŞ T.A.Ş Verileri için Parametre Tahmin Değerleri _	[93]
Tablo-4.37 , İzmir ÇİMENTAŞ T.A.Ş Verileri için Sapan Değer Analizi Son.____	[93]
Tablo-4.38 , İzmir ÇİMENTAŞ T.A.Ş Verileri için Sapan Değer Analizi Son.____	[93]
Tablo-4.39 , Gıda Sektöründe Unların Su Tutma Yüzde Değerleri_____	[94]
Tablo-4.40 , Gıda Sekt. Unların Su Tutma Yüzde Değ. için Değişik İterasyon...____	[94]
Tablo-4.41 , Gıda Sekt. Unların Su Tutma Yüzde Değ. için Param. Tahmin Değ____	[95]
Tablo-4.42 , Gıda Sekt. Unların Su Tutma Yüzde Değ. için Sapan Değ. Sonuçları_	[95]
Tablo-4.43 , Gıda Sekt. Unların Su Tutma Yüzde Değ. için Sapan Değ. Sonuçları_	[95]
Grafik-4.1 , 64. Pozisyon için Kuşku Gözlem Serisi_____	[89]
Grafik-4.2 , 64. Pozisyon için Kuşku Gözlem Serisi_____	[90]

KISALTMALAR

- ARIMA : **Auto-Regressive-Integrated-Moving-Average.**
- ARMA : **Auto-Regressive-Moving-Average.**
- AR : **Auto-Regressive.**
- MA : **Moving-Average.**
- HKT : **Hata Kareler Toplamı.**
- HKO : **Hata Kareler Ortalaması.**
- VA : **Varyans Analizi.**
- Bkz : **Bakınız.**
- LS : **Least Squares.**
- M-H : **Method of Huber-type "down-weighting" .**
- M-B : **Method of Bisquare-type "down-weighting" .**
- GM-H : **Generalized Method of Huber-type "down-weighting" .**
- GM-B : **Generalized Method of Bisquare-type "down-weighting" .**
- B : **Backward Shift Operator.**
- F : **Forward Shift Opeartor.**
- IO : **Innovational Outlier.**
- AO : **Additive Outlier.**
- max : **Maksimum.**
- min : **Minimum.**
- C : **Critical value.**
- D.K : **Değişim Katsayısı.**
- S.D : **Serbestlik Derecesi.**
- K.T : **Kareler Toplamı.**
- K.O : **Kareler Ortalaması.**
- F : **F-test değeri.**
- p : **(probability) Olasılık Değeri.**
- T : **Sapan değerin bulunduğu pozisyon.**

BİRİNCİ BÖLÜM

1. GİRİŞ

Bir deyişe göre, "Sayılar her zaman doğruyu söyler". Bu durumda sayılar çok büyük önem kazanmaktadır. Sayıların söylediği veya göstermek istediği gerçekleri ortaya çıkarabilmek, sayıların objektif bir şekilde toplanması, incelenmesi ve değerlendirilmesi ile mümkündür. Doğaldır ki ne kadar çok sayı bir araya getirilirse, rassal nedenlerle ortaya çıkabilecek sapmalar azalacak ve genel eğilim bütün açıklığı ile ortaya çıkacaktır. Önceleri matematiğin bir alt disiplini olarak ortaya çıkan istatistik, çok sayıya sahip olmayı "Büyük Sayılar Kanunu" ile birleştirmek suretiyle kısa sürede büyük gelişme göstermiş olup toplum, devlet, firma ve birey için vazgeçilmez bir analiz aracı haline gelmiştir (Aytaç, 1994).

Bir olaya ilişkin verilen karar veya yapılan bir iş, yarımın ne getireceği hakkında bazı beklentilere dayanır. Gelecek hakkında bu gibi beklentiler, tahmin işlemleri veya yöntemlerin sonuçlarıyla daha anlamlı bir boyut kazanır. Tahmin yöntemleri, basit yöneticilik tekniklerinden karmaşık ve zaman alıcı veri analizlerine kadar çeşitlilik gösterir. Dolaylı veya dolaysız tahminler, rasyonel kararların alınmasında ve yürürlüğe konulmasında mutlaka gereklidir. Buradan hareketle belirli bir konu ile ilgili olarak verilerin uygun şekilde toplanması, açık ve anlaşılır bir tarzda ifade edilmesi büyük önem taşımakta, genel anlamda istatistiğin önemli bir aşamasını meydana getirmektedir. Ancak, genel anlamda istatistiğin diğer aşamaları olan, veri analizi ve istatistik karar teorisi, önemi açısından ön plana geçmiştir. Özellikle sınırlı bilgilerden sonuç çıkarmak ve bu sonuca göre karar vermenin gerekli olduğu durumlarda istatistik metodların uygulanması, çeşitli alanlarda ilerleme imkanı yaratmıştır. İstatistik analizde tahmin aracı olarak kullanılan yöntemlerden biri de zaman serisi analizidir.

Ülkemizde özellikle ekonomik eğilim ve tahminleri belirlemeye yönelik yapılan ekonometrik çalışmaların önemli bir bölümü zaman serilerinin kullanımı ile yapılmaktadır. Milli Gelir, Üretim, Fiyatlar, İstihdam gibi reel ekonominin temelini

oluşturan unsurlar için zaman serisi analizleri yoğun bir şekilde kullanılmaktadır. Ayrıca kur tahminleri, bütçe ile ilgili tahminlemeler, hisse senedi eğilimleri, enflasyon tahminleri ve daha bir çok tahminleme işlemi zaman serisi analizleri ile yapılmaya başlanmıştır. Sürekli üretim süreçlerinde hata tespit edilmesine yönelik kullanılan kalite kontrol aşamalarında, hava tahmini verilerinde, düzeltme yöntemlerinin etkinliklerini artırmak amacıyla uç değerlerin belirlenmesi işlemlerinde, astronomik gözlem değerlerinin yorumlanmasında ve daha bir çok alanda sapan değer analizi çalışmaları yapılmaktadır.

Bilgisayarın etkin bir araç olarak bilim ve iş dünyasına girmesi, zaman serileri ve sapan değerlerin analiz aracı olarak kullanılmasını kaçınılmaz kılmıştır. Bu sayede sapan değer analizleri, bütün istatistik analizler için temel analiz özelliği kazanmıştır. Bir başka deyişle sapan değer, "Örneklemedeki diğer gözlemlerden farklı olan ve kestirim sonuçlarını büyük ölçüde etkileyen az sayıda gözlem değeri" olarak ifade edilir. Sapan değerleri tespit etmek amacıyla bir çok yöntem geliştirilmiştir. Simülasyon çalışmaları ve denemeler, bu yöntemlerden en etkin olan, *Chang, Tiao ve Chen (1988)* tarafından geliştirilerek ardışık yöntem adını verdikleri yöntem olduğunu ortaya koymuştur.

Bu çalışmanın kapsamı içerisinde sapan değerler konusuna ilişkin kaynak taramaları verilerek, zaman serileri ve sapan değerler konusunda yapılan veya yapılmakta olan çalışmalar hakkında bilgi verilmiştir. İkinci bölümde; bilinen zaman serileri tahmin modellerinin tanıtımı ve karşılaştırmaları yapılmış, ayrıca Box-Jenkins (B.J) tahmin modelleri olarak bilinen stokastik zaman serisi modelleri (AR, MA, ARMA, ARIMA) için özet bir tanıtım verilmiştir. Üçüncü bölümde; sapan değerler konusunda tarihi gelişim dikkate alınarak, basit anlamda sapan değer tespit yöntemleri ve zaman serilerinde ortaya çıkması muhtemel sapan değerler tanıtılmıştır. Ayrıca zaman serilerinde sapan değerlere ilişkin test yöntemleri, test yöntemlerinin performansı ve bu yöntemlerden etkin olan Ardışık Yöntem'in teorik yapısı tanımlanmıştır. Dördüncü bölümde sapan değerler tespit edilirken simülasyon çalışmaları ile elde edilen veriler kullanılarak varyans analizi yöntemi ile;

- Model farklılığı (AR, MA),
- Sapan değer büyüklükleri (3σ , 5σ),
- Seri genişliği (50, 100, 150),
- Ölçüt değer duyarlılığı (3.00, 3.50, 4.00),
- Sapan değer türü (AO, IO),
- Sapan değer sayısı (1, 2, 3),
- Sapan değer tarama yöntemleri (LS, M-H, M-B, GM-H, GM-B, ARDIŞIK)

gibi ana faktör ana etkilerinin sapan değer tespitinde ne derece etkili olduğu araştırılmıştır. Bununla birlikte;

- Kontrol amaçlı (kalite kontrol amaçlı süreçler için) Box-Jenkins A serisi,
- Veri setlerini düzeltme amaçlı, ÇİMENTAŞ T. A. Ş 1980-87 aylık çimento ihracat verileri,
- Parametre tahminlerinde etkinlik artırma amacına yönelik olarak ta, Gıda sektöründe unların yüzde su tutma kapasitesini belirlemek amacıyla elde edilmiş değerler kullanılarak sapan değer analizleri yapılmıştır.

Beşinci bölümde ise, teorik bulgularla birlikte elde edilen üç farklı veri seti için sonuçlar verilmiştir.

1.1. Kaynak Taramaları

Çalışmada incelenen sapan değer kavramı, genel anlamda ilk kez 1838 yılında Alman astronomlar okulunda kullanılmaya başlanmıştır. Daha sonraları istatistik biliminin gelişmesi ve araştırma çalışmalarının artmasıyla doğrusal modeller, regresyon modelleri, çok değişkenli istatistik analiz, deney düzenleme ve son olarak zaman serilerinde sapan değer çalışmaları yapılmaya başlanmıştır (*Tatlıdil, 1981*).

Zaman serilerinde sapan değerler kavramı ilk kez *Fox(1972)* tarafından çalışılmıştır. *Fox*, otoregresif modellerde sapan değerleri ortaya çıkarmak amacıyla olabilirlik oran kriteri adını verdiği bir test yöntemi geliştirmiş ve ortaya çıkan sapan değerleri birinci ve ikinci tip sapan değerler biçiminde tanımlamıştır. Ayrıca *Fox*, güç fonksiyonları üzerine çalışmıştır (*Fox, 1972 ve Hawkins, 1981*).

Daha sonraları bir çok araştırmacı, *Fox'* un yaptığı çalışmaları daha da geliştirerek, Bütün ARIMA modelleri kapsayacak şekilde, çoklu sapan değerleri taramaya yönelik yöntemler geliştirmiştir. Bu yöntemi ardışık yöntem olarak tanımlamışlardır. *Hillmer (1983), Tsay (1986), Pena (1987), Abraham ve Yatawara (1988), Bruce ve Martin (1989)* Ardışık Yöntemin teorik yapısı üzerinde çalışmalar yapmışlardır.

Bununla birlikte *Abraham ve Yatawara (1988)* lagrange çarpanlar metodu veya skor tabanlı sapan değer testleri üzerine çalışmalar yapmışlardır. *Pena (1987), Abraham ve Chuang (1989), Bruce ve Martin (1989)* sapan değer taramada, sapan olan gözlemin silinmesi esasına dayalı testler ile birlikte zaman serilerinde etkili gözlemler üzerine çalışmalar yapmışlardır (*Ljung, 1993*).

Diğer yandan *Harrison ve Stevens (1976), Smith ve West (1983), West, Harrigon ve Migon (1985)* ayrıca *West (1986)*, zaman serilerinde sapan değerleri belirlemek için sıralı tarama yöntemleri üzerinde çalışmışlardır (*Tsay, 1988*).

Zaman serilerinin analizi için kullanılan sıralı tarama yöntemleri, doğrusal regresyon problemlerinin çözümünü kapsayacak şekilde geliştirilmiştir. Bu amaçla, *Mararsinghe (1985)*, *Kianiard ve Swallow (1990)* tarafından sıralı strateji testi (Sequential Strategy Test) adıyla bir yöntem geliştirilmiştir (*Pena ve Yohai, 1995*).

Ayrıca, *Jones (1980)*, *Ljung (1982, 1989)*, *Harvey ve Pierce (1984)*, *Kohn ve Ansley (1986)*, *Wincek ve Reinsel (1986)* ve diğerleri, gözlem değerlerinden bazılarının kayıp olması durumunda, sapan değerlerin tespit edilmesine yönelik olarak olasılık oran kriterleri geliştirmişlerdir (*Ljung, 1993*).

Zaman serilerinde sapan değerlerin etkisi üzerine yapılan çalışmalara *Box ve Tiao* da katılmıştır *Box ve Tiao* yaklaşımı, *Chen ve Tiao (1986)* tarafından, *Chang ve Tiao (1983)*'nun ardışık yöntemi kullanılarak genelleştirilmiştir. Son zamanlarda ise *Tsay* buna yeni eklemeler yapmıştır. Birinci dereceden AR modellerde sapan değer etkileri ise *Whichern, Miller ve Hsu (1976)* tarafından tahmin edilmiştir. Bu araştırmacılar varyans değişimlerinin parametre tahminleri üzerindeki etkisini incelemiştir (*Tsay, 1988*).

Zaman serilerinde sapan değerlerin tespit edilmesine yönelik çalışmalardan biri de *Denby ve Martin (1979)* tarafından geliştirilen yöntemdir. Bu yöntem *Martin ve Yohai (1985)* tarafından da çalışılmıştır. Ancak *Chang ve Tiao (1983)* söz konusu yöntemin ikinci tip sapan değerlerin teşhis edilmesinde etkin olmadığını ortaya koymuştur (*Tsay, 1988*).

Bruce ve Martin (1989) birinci tip sapan değerlerin tahminlenmesine yönelik çalışmalarında, sapan değerlerin sadece bulunduğu pozisyonu değil, yakın pozisyonları da etkilediğini ortaya koymuşlardır.

Chang (1988), yanlış belirlenen sapan değer tiplerinin, test yöntemlerinde etkinlik kaybına neden olduğunu ortaya koymuştur (*Muirhead, 1986*).

Sapan değerleri tespit etmede kritik değer olarak kullanılan C (ölçüt değer) değerlerini saptamak amacıyla Monte-Carlo simülasyonları kullanılmıştır. Ayrıca, *Berman*

(1964) tarafından bulunan ve bu amaçla kullanılan asimptotik sonuçların da kullanılabilir olduğu yine *Berman* tarafından ıspat edilmiştir (*Ljung* 1993).

Sapan değerler kavramı, bir çok arařtırmacı tarafından farklı biçimlerde tanımlanmıştır. Mesela sapan değer *Bross* (1961) tarafından, "Görünümü geri kalan gözlem değerlerinin görünümüne uymayan bir veya bir kaç gözlem" olarak tanımlanmaktadır. *Stefansky* (1972), "kestirim değerlerinden çok farklı olan gözlem değerleri" şeklinde bir tanım vermiştir. Aynı tanım, "örneklemdaki diğer gözlemlerden farklı olan ve kestirim sonuçlarını büyük ölçüde etkileyen az sayıda gözlem değeri" şeklinde de verilmiştir. Bu ve buna benzer bir çok tanım verilebilir. Ancak bir gözlemin sapan olabilmesi değer olarak büyük yada küçük olmasıyla ilişkilendirilemez. Gerçek model denklemi elde edildiğinde büyük hata kestirimine sahip olan gözlem veya gözlemlerin sapan olmaları olanaklıdır. Bu durumu gözönünde bulundurarak, "gözlem ve tahmin değerleri arasındaki farkın en büyük olduğu gözlem, sapan gözlem olabilir" denilebilir (*Kurt*, 1990).

Sapan değerler; model yanlışlığı, gerekli bazı dönüşümlerin yapılmamış olması, ölçüm, tartım ya da kaydetme hataları gibi çeşitli nedenlerden kaynaklanabilmektedir. Örneklemdaki bazı gözlemlerin, diğer gözlemlerden farklı bir etkenin ya da etkenlerin etkisinde kalmış olması da sapan değer oluşturabilmektedir. Bu değerler sadece doğal rastgelelik sonucunda da ortaya çıkabilmektedir (*Tatlıdil*, 1981).

Zaman serilerinde sapan değerleri ortaya çıkarmak amacıyla bir çok yöntem geliştirilmiştir. Simülasyon çalışmaları ve denemeler, bu yöntemlerden en etkili olan *Chang, Tiao ve Chen* (1988) tarafından geliştirilen ve Ardışık Yöntem adını verdikleri yöntem olduğunu ortaya koymuştur. Söz konusu yöntem, diğer yöntemlere göre uygulanması kolay ve pratik olmasına rağmen karmaşık matematiksel işlemleri gerektirmektedir (*Chang, Tiao ve Chen*, 1988).

Kalite kontrolde, sürecin kontrol altında olup olmadığını belirlemek amacıyla, süreçten elde edilen gözlemlerin bir zaman serisi oluşturduğu düşünülerek serinin uyum gösterdiği parametre tahminleri elde edilmekte, daha sonra seride ortaya çıkması olası iki

tip sapan deęer belirlenmektedir. Sapan deęer tiplerine gre problem; girdi, ıktı ya da sre problemi olarak sınıflandırılmaktadır (*Bayhan, 1992*).

Amerika'da finansal politikaları tahminlemek amacıyla *Giorgio Canerella* ve *Stephen K.Polard* tarafından yapılan alıřma, *Akaike (1973)*, *Schwarz (1978)* ve *Hannan (1980)* tarafından geliřtirilen, ARIMA modellerin optimal zelliklerini yansıtan lt alıřma esas alınarak yapılmıřtır (*Pena, 1987*).

Benzer olarak ARIMA ve ekonometrik modeller arası baęlantı kullanılarak, ARIMA modeller zerine alıřmalar yapılmıřtır. Bu alıřmalar Amerika'da *Nelson (1973)*, *Granger ve Newbold (1977)*, *Nefti (1982)*, *Harvey ve Tood (1983)* tarafından yapılmıřtır. Daha sonraları *Berkman (1980)*, *Simpson (1980)* ve *Hafer (1980)* ARIMA modeller zerine yapılmıř yeni tanımlamalarla tahmin edilmiř parametre deęerleri zerine ayrıntılı alıřmalar yapmıřlardır. Ayrıca Amerika'da *Froyen (1974)*, *Barro (1978)*, *Niskanen (1978)*, *Levy (1981)*, *Sims (1980)*, *Mishkin (1982)*, *Illmakunnas ve Tsurumi (1985)* ve daha bir ok arařtırmacı, ekonomik ve ekonometrik amalı alıřmalar yapmıřlardır (*Pena, 1987*).

Ekonomik amalı alıřmalar yanında, meteorolojik alıřmalarda, saęlık bilimlerinde, hava ve evre kirlilięi ile ilgili alıřmalarda ve szkonusu kirliliklerin yol atıęı hastalıklar ile ilgili alıřmalarda artan oranda ikili zaman serileri (binary time series) ile alıřılmaktadır. Bu amala Amerika' da saęlık ve evre koruma ajansı tarafından alıřmalar yapılmaktadır (*Fitzmaurice ve Lipsitz, 1995*).

Zaman serileri ile ilgili yapılan alıřmalardan biri de kısmi otokorelasyona dayalı hata kareler toplamı iin iyi uyum testleri ile ilgili alıřmalardır. Bu alıřmalardan biri de *Anna Clara Monti* tarafından yapılmıřtır. Bu amala geliřtirilmiř bir ok test bulunmaktadır. Szkonusu testler *Quenouille (1947,1949)*, *Bartlett (1954)*, *Box ve Pierce (1970)*, *Ljung ve Box (1979)*, *Ansley ve Newbold (1979)* ve *Godfrey (1979)* tarafından geliřtirilmiřtir (*Monti, 1994*).

Dokuz Eylül Üniversitesi'nden Dr. Abdul Sattar Rashid Salim Al-Khalidi, su kaynaklarının kirlilik ölçümlerinde analiz modelini, tahminlenebilen dışsal etkiler ve açıklanamayan hata terimlerinin doğrusal bir kombinasyonu olarak ifade etmiştir. Kirliliğe etki eden bileşenler için parametreleri belirlenmiş ve bileşenlerin etkileri tahminlenerek, çalışma sonuçlandırılmıştır (*Al-Khalidi, 1997*).

Hacettepe Üniversitesi'nden Dr. Hüseyin Tatlıdil, hava tahmini verilerini kullanarak doğrusal regresyon modellerindeki kuşkulu gözlem değerlerinin testinde kullanılan yöntemlerin karşılaştırılması, bu yöntemlerden daha kullanışlı yeni bir yöntemin geliştirilmesi, doğrusal regresyon modelleri için önerilen bu yöntemin doğrusal olmayan regresyon modellerinde ve çok değişkenli veriler üzerinde kullanılıp kullanılmayacağı üzerine bir doktora çalışması yapmıştır (*Tatlıdil, 1981*).

Dokuz Eylül Üniversitesi'nden Dr. Ali Şen, Sapan ve etkili gözlemlerin bulunmasında kullanılan istatistikler başlığı altında, tam ranklı doğrusal regresyon modellerinde parametre vektörünün en küçük kareler tahminlerinin belirlenmesine her veri noktasının katkısının anlaşılabilmesi için güven elipsoidlerine dayanan bir istatistik geliştiren Cook istatistiğini, bir'den fazla veri noktasının etkisini inceleyebilen AP (Andrews-Pregibon) istatistiğini ve son olarak bu iki istatistiği ele alarak bir araştırma çalışması yapmıştır. Bu çalışmada Cook belirli bir modele göre tek bir gözlem etkisini incelerken AP, bir veya daha fazla gözlemin sapan olup olmadığına ait bilgi içeren bir yöntemi ifade etmektedir. Ayrıca Cook istatistiği, AP istatistiğinin iki bileşeni ile birlikte gözlemlerin faktör uzayındaki uzaklıklarına ait bilgilerin dikkate alınması gerektiğini ifade etmektedir (*Şen, 1993*).

Dokuz Eylül Üniversitesi'nden Dr. Serdar Kurt, çok etkenli deneylerde tek sapan değer çözümlemesi üzerine bir seminer çalışması yapmıştır. Bu çalışmada sapan değer, "kestirim değerlerinden çok farklı olan gözlem değeri olarak tanımlanmıştır". Buna göre bir gözlem değerinin sapan olup olmadığının göstergesi hata kestirim değeridir. Araştırmacı elde ettiği hata kestirimlerinden herhangi birinin değerini diğerlerinden oldukça farklı

olarak görürse, söz konusu hata terimine ilişkin gözlem değerini kuşkulu gözlem olarak tanımlayıp sapan değer testi uygulamıştır (*Kurt, 1990*).

Ayrıca, sağlık bilimlerinde ilaç tedavilerinin ani etkileri, bunlara etki eden bileşenlerin faktörler ve etkileşimleri üzerine çalışmalar yapılmıştır. Bilgisayar Bilimlerinde ise, bir bilgisayar sisteminde ortaya çıkan arıza durumlar ve veri tabanları üzerinde kişi veya sistem ortamının olumsuzluğundan ortaya çıkabilecek hata durumları, sapan değerlerle ilişkilendirilerek, tespitler yapılmakta ve çözüm önerileri sunulmaktadır.



İKİNCİ BÖLÜM

2. ZAMAN SERİLERİ

Zamanın fonksiyonu olarak elde edilmiş verilerin oluşturduğu seriler zaman serileri olarak bilinir. Geleceğe dönük tahminler yapmanın en önemli yollarından biri olan zaman serilerinde birbirlerini izleyen gözlem değerlerinde zamana göre bağımlılık söz konusudur. Seriyeye uyumu sağlanacak bir fonksiyon ile gelecek gözlemler tahminlenebilir. Zaman serileri uygulama alanları içerisinde ekonomik zaman serileri en başta gelen konular içerisinde yer almaktadır. Milli gelir, Üretim, Tüketim, Satışlar, İhracat, İthalat ve bunların zamana göre değişiminin incelenmesi, gelecek için tahminlenmesi ayrıca meteorolojik, sosyolojik, demografik, sağlık ve kalite kontrolde zaman serisi analizlerinden oldukça yoğun bir şekilde yararlanılmaktadır.

İncelenen bir zaman serisi trend, konjonktürel dalgalanma, devresel hareketler ve zaman gibi faktörleri içerebilir. Faktörler arasında zaman dışındaki diğer etkiler için hesaplanan endeksler ile gözlemlerden söz konusu etkiler yok edilir. Bu işlem sonucunda zaman serisi analizleri uygulanabilir. Bu durumda zamanın fonksiyonu olarak elde edilen serilere zaman serisi analizleri uygulanabilir. Eğer bir zaman serisi kesin olarak tahminlenebiliyorsa "deterministik zaman serisi" olarak isimlendirilir. Zaman serilerinde asıl amaç, gözlemlerle elde edilen değerlerin olasılık kuralları içerisinde saptanmasıdır. Bu tip serilere "stokastik zaman serileri" denir.

İncelenen gözlemler belirli zaman aralıklarında elde ediliyor ise, bu tür serilere "kesikli zaman serileri" adı verilir. Bu tür seriler genellikle eşit zaman aralıkları ile elde edilen verilerden oluşur. Başka bir deyişle ele alınan gözlemler zaman içerisinde belirli periyotlara göre belirlenmiş ise bu tip serilere kesikli zaman serileri denir. Günlük, haftalık, aylık ve yıllık bazda elde edilen gözlemlerden oluşan seriler kesikli zaman serilerine örnek verilebilir.

Zaman serisi analizinin yapılabilmesi için, parametre tahminlerinde yanlılığa sebep olan, bir başka deyişle ileri sürülen durağanlık gibi varsayımların göz önünde tutulmamasına yol açan trend, konjonktürel dalgalanma, mevsimlik etki ve rasgele değişkenliklerin kontrol edilmesi gerekmektedir.

Bir zaman serisinin uzun dönemde bir yöne doğru gösterdiği eğilime trend denir. Nispeten mantıksal ve basit fonksiyonlarla ifade edilen düzenlilikler güvenli bir biçimde ileriye uzatılabilir. Bu durum; nüfus artışı, göçler, üretim artışı gibi artışlarla ilgilidir. Örneğin hava yolu taşımacılığının büyümesi, nüfus artışı ve taşıma sektöründe havayolu payının artışı gibi iki nedene bağlı olabilir. Bu durumda büyüme nispeten hızlıdır. Bu gibi uzun dönem değişimine dönemsel trend denir. Örneğin bir ülke veya bölgede nüfus artışının trendi elde edilmek isteniyorsa 20, 30 veya daha uzun dönemlik bir veri setinin incelenmesi gerekli iken bir bakteri grubunun büyüme hızı 24 saatlik bir gözlemden sonra tamamlanabilir.

Mal ve hizmetlerin üretim, satış, tüketim ve fiyatlarında iklime bağımlılıktan ötürü bazı değişimler söz konusu olabilir. Örneğin tarımsal ürünlerin fiyatlarında mevsimler bazında görülen değişimler, yolcu taşıma sayılarında meydana gelen farklılıklar, bayram satışlarında meydana gelen patlama, yaz aylarında inşaat sektöründe görülen canlılık, yine yaz aylarında su tüketimi ve benzeri faaliyetlerde bariz bir artış söz konusu olur. Bunun gibi bir yıl içerisinde aylık olarak elde edilmiş gözlemlerde inişler ve çıkışlar görülür. Bu tip hareketler her yıl ortaya çıkmaktadır. Bu tür mevsimsel etkinin varlığı aylık periyotlarda görülebilmektedir. Öte yandan yaz ve kış tatilleri, günlük çalışma saatleri ve daha benzer bir çok sebepten kaynaklanan faktörler, periyodikliği doğurabilmektedir.

Gerçek gözlemler etrafında 3-5 yılda ya da daha uzun zaman periyotlarında görülen değişimler, konjonktürel dalgalanmalar olarak bilinir. Daha açık bir ifade ile Üretim, Satışlar, Gelirler ve Yatırımlar gibi iktisadi olaylarda bir süre gelişme görülür. Bu yükselmenin maksimuma çıktığı anda bir ekonomik kriz başlar, bu düşüşten bir süre sonra faaliyetler tekrar canlanır ve aynı olaylar tekrarlanır. Bu safhalar ilgili zaman periyotları içerisinde tekrarlanır. Zamanın fonksiyonu olarak elde edilen değerler

yükselerek seyrediyor iken konjonktürün de yükselmesi değerlerin artışını hızlandırır. Aksine konjonktörün düşmesi yükselme eğilimini hafifletebilir.

Yukarıda açıklanan dalgalanmalar dışında ekonomik olaylara etki eden bazı hareketler söz konusu olabilir. Bunlar; savaş, deprem, grev, sel, don, dolu, siyasi karışıklıklar ve dış ticaret güçlükleri gibi ne zaman nasıl tekrarlanacağı belli olmayan hareketlerdir. Bu hareketler bir çok ekonomik değeri büyük ölçüde etkileyebilirler.

O halde belirli bir olay hakkında gözlemlerin kronolojik olarak sıralanmasından doğan bir zaman serisi kesin olmamakla birlikte dört ayrı güç tarafından etkilenir. Bu güçler; trend, mevsimlik değişimler, devresel değişimler ve sayılan üç etkiden kaynaklanmayan değişimlerden kaynaklanan ve hata değişimleri olarak bilinen değişimlerdir (*Aslan, 1990*).

Zaman serileri başlıca;

- Zaman serisini unsurlarına ayırmak,
 - Zaman serileri arasındaki ilişkiyi açıklamak,
 - Süreç kontrolü,
 - İleriye dönük tahminler
- yapmak amacıyla analiz aracı olarak kullanılırlar.

Zaman serilerinin ileri dönük tahmin yöntemleri iki grupta toplanabilir. Bunlar, çok değişkenli ve tek değişkenli zaman serileri ile ilgili tahmin yöntemleridir. Çok değişkenli zaman serileri iki veya daha fazla zaman serisi arasındaki sebep-sonuç ilişkisini tanımlayan ve daha sonra tahmin ve kontrol amacıyla kullanılan serilerdir. Üzerinde durulan değişken ile bu değişkeni açıklayan diğer değişkenler arasında mantıksal ilişkiler varsa ve bu değişkenlerin zaman aralıklarıyla aldığı sayısal değerler mevcut ise bir ilişki modeli kurulur. Kaynaklarda dönüşüm fonksiyon modelleri, dinamik regresyon modelleri ve çok değişkenli zaman serileri analizi olarak bilinen yöntemler örnek verilebilir.

Çok değişkenli zaman serileriyle ilgili modeller, tahmin sistemiyle ilgili herşeyin bilindiğini dikkate alır ve birbiriyle ilişkili olayların tahmin edilmesini sağlar. İlişki modellerine dayanarak yapılan tahminlerin hatası düşük olabilir. Ancak tahmin sistemiyle ilgili herşeyin bilinmesi çoğu zaman mümkün olmayabilir veya mümkün olsa bile analiz için uygun olmayabilir. Bu durumu ekonomide bir örnekle ifade etmek gerekirse; milli gelir değişkeni çok sayıda ekonomik değişken için açıklayıcı değişken durumundadır. Ancak yıllık değerleri bilinen bir değişkenin aylık değerlerini istatistiksel yöntemlerle belirlemek mümkün olsa da belirlenen değerler sağlıklı olabilir. Bu nedenle aylık verileri esas alan ekonometrik çalışmalarda kullanılan ilişki modellerinde sağlıklı olmayan verilerin kullanılması ile elde edilen bulguların güvenilirliği düşük olabilir. Bu nedenlerden dolayı zaman serilerinin ileriye dönük tahmininde tek değişkenli zaman serileriyle ilgili tahmin yöntemleri yaygın bir şekilde kullanılmaktadır.

Tek değişkenli zaman serileri ile ilgili tahmin yöntemleri, zamana bağlı tek bir değişkene ait verilerin mevcut olması durumunda kullanılan ve sadece ileriye dönük tahmin yapmaya imkan veren istatistiksel yöntemlerdir. Bu yöntemler zaman serilerinin bu günkü ve geçmiş dönem gözlem değerlerini kullanarak, gelecek dönem tahmin değerlerinin elde edilmesini sağlarlar.

Bu grupta toplanan yöntemlerin dayandığı varsayımlar aşağıda sunulmuştur;

- Bir zaman serisinde mevcut olan zaman serisi unsurlarının gelecek dönemde de aynı kalacağı kabul edilir. Bu varsayım nedeniyle geçmiş dönem gözlem değerlerine dayanarak gelecek dönem tahmin değerleri elde edilir.
- Bu yöntemler, zaman serisini meydana getiren unsurları birbirlerinden ve tesadüfi unsurlardan ayırmak suretiyle serinin gelecekte alabileceği değeri tahmin etmeyi amaçlar.
- Bu yöntemler, eşit zaman aralıklarıyla elde edilen gözlem değerlerinden meydana gelen kesikli zaman serilerine uygulanır.

Tek deęişkenli zaman serileriyle ilgili temel analiz yöntemleri;

- Trend Analizi,
- Hareketli Ortalamalar Analizi,
- Üssel Düzeltme Analizi,
- Uyarlayıcı Arındırma Analizi.

Geliştirilen analiz yöntemleri arasında en çok bilineni trend analizi yöntemidir. Hesaplanması ve anlaşılması kolay olan bu yöntem, günümüzde orta ve uzun dönem tahmin amacıyla yoğun bir şekilde kullanılan sayısal tahmin yöntemlerinden biridir. Trend analizinin esası, zamana baęlı herhangi bir olaya ait deęerlerin serpm diyagramında göstermiş oldukları serilmeye uygun matematiksel bir fonksiyon belirlemek ve bu fonksiyonla ilgili olayın zamana göre nasıl bir eğilim gösterdiğini tespit etmektir. Trend analizi; biri açıklayıcı zaman, dięeri açıklanan iki deęişken ile kurulan matematiksel eşitlik üzerinde yapılan bir analiz yöntemi olarak ifade edilir. Yöntemde sakınca olarak görülebilecek unsurlardan birincisi, tahmin işleminde sadece iki deęişkenin dikkate alınmasıdır. Oysa bir oluşumun meydana gelmesinde etken olan çok sayıda faktör bulunur.

Hareketli ortalamalar analizinin esası, bir zaman serisindeki gözlem deęerlerini belirli büyüklükteki kümeler halinde toplamak, her küme için aritmetik ortalama hesaplamak ve bu ortalamayı, ait olduęu kümenin en yeni terimini izleyen terimin tahmin deęeri olarak kabul etmektir. Hareketli ortalamalar tahmin yöntemi gözlem deęerlerinin oluşumunda rasgelelięin yüksek olduęu, buna karşılık birbirini izleyen gözlem deęerleri arasındaki otokorelasyonun düşük olduęu zaman serilerinde uygulanır. Kısa dönem tahmin amacıyla kullanılabilir olan bu yöntemin uygulanabilmesi için çok sayıda gözlem deęerine gereksinim vardır. Hareketli ortalama yöntemine yapılan en ciddi eleştiri, bu yöntemin sadece hareketli ortalama dönemindeki tarihi verilere eşit aęırlık vermesi, eski dönemleri bütünüyle görmezden gelmesidir.

Üssel düzeltme yöntemi, mevsim ve trend unsuru içermeyen basit formdaki zaman serileri için uygulanmıştır. Bu yöntemlerin her biri kendinden önce geliştirilmiş olan yöntemlerin dezavantajlarını avantaja dönüştürmeyi amaç edinmiştir. Bu nedenle üssel düzeltme yöntemlerinin (Extrapolation of trend curves, Exponential smooting, The Holt-Winters Forecasting ve benzerleri gibi) bazıları diğerlerine oranla daha çok yönlü, bazıları hesaplama açısından karmaşıktır. Bazılarının analizi için ise oldukça fazla bilgisayar zamanına ihtiyaç duyulur. Üssel düzeltme yöntemleri, temel özellik olarak hareketli ortalama tahmin yöntemine benzemesine karşın, zaman serilerinin tüm gözlem değerlerini gözönünde bulundurdıkları ve seri değerlerine bu günkü dönemden uzaklıklara göre azalarak tartı verdikleri için hareketli ortalama yönteminden ayrılırlar.

Uyarlayıcı arındırma yönteminde ise, zamana bağlı bir olayla ilgili tahmin modeli belirlendikten sonra, bu olayı meydana getiren unsurlarda meydana gelebilecek değişiklikleri yeniden bir tahmin modeli belirlemeye gerek bırakmadan doğrudan tahmin değerlerine yansıtma imkanı olan bir yöntemdir. Bu modeller tahmin işleminde de araştırmacının müdahalesini minimum düzeye indirir. Uyarlayıcı arındırma tahmin yöntemine göre herhangi bir gelecek dönemin tahmin değeri, hareketli ortalamalar ve üsse düzeltme yöntemlerinde olduğu gibi geçmiş dönem gözlem değerlerinin toplamı alınarak elde edilir. Uyarlayıcı arındırma yöntemi, hareketli ortalama ve üssel düzeltme yöntemleri gibi kısa dönem tahmin amacıyla kullanılır.

Yukarıda kısaca tanımlanan ve karşılaştırılan dört farklı analiz yöntemi, tek değişkenli zaman serilerinin tahminine yönelik olarak kullanılan değişik yöntemlerdir. Söz konusu yöntemler, zaman serisi gözlem değerlerinin karakteristiklerine bağlı olarak değişik üstünlük ve avantajlar sağlarlar. Bu yöntemler zaman serilerinin ardışık gözlem değerleri arasında var olan bağımlılığı dikkate almazlar.

2.1. Box-Jenkins (B.J) Tahmin Modelleri

Yukarıda tanımlanan ve karşılaştırılan yöntemlerde bir zaman serisi için hesaplanan ortalama değer, zamanın deterministik bir fonksiyon olduğu varsayımına dayandırılarak yapılır ve bu ortalamaya belirli bir dönemin hata terimi ilave edilerek o dönemin gözlem değeri elde edilir. Oysa zaman serilerinin çoğunda ardışık gözlem değerleri birbirine bağımlıdır. Bu durumda trend analizi için hareketli ortalama ve üssel düzeltme yöntemleri uygun değildir, çünkü bu yöntemler bağımlılık avantajını kullanmazlar. Halbuki zaman serilerinde bağımlılık yapısını çok etkin bir şekilde kullanan Box-Jenkins tahmin modelleri olarak da bilinen ARIMA modeller bulunmaktadır.

Box-Jenkins tahmin modelleri tek değişkenli zaman serilerinin ileriye dönük tahminleri için kullanılan başarılı bir tahminleme yöntemidir. (B.J) grubu modeller, zamana bağlı olayların rasgele karakterde olması ve bu olaylarla ilgili zaman serilerinin stokastik süreç olduğu varsayımına dayanarak son 30-40 yılda geliştirilmişlerdir. Ayrıca bu modellerde rasgele değişkenin zaman içinde ardışık olarak aldığı değerler arasında mevcut olan otokorelasyon en etkili bir şekilde dikkate alınır. Bu nedenlerden dolayı sözkonusu modellere stokastik modeller adı verilmektedir.

Bu modeller için yapılan en önemli varsayımlar gözlem değerlerinin eşit zaman aralıklarıyla elde edilmesi, bir başka deyişle kesikli ve durağan olmasıdır. Zaman serilerinde duraganlık; trend, mevsimsel dalgalanma ve benzeri etkilerin seriden silinmesi veya tahmin değerlerini önemli derecede etkilemeyecek oranda kalması anlamına gelmektedir. Serinin durağanlığı üzerine konmuş varsayım çoğu zaman bozulmaktadır. Durağan olmayan zaman serilerinin ileriye dönük tahmininde Box-Jenkins yönteminin uygulanabilmesi için önce durağanlığı bozan söz konusu etkenlerin ortadan kaldırılması gerekir.

2.1.1. AR Modeller

Bu modeller; bir zaman serisinin herhangi bir dönemdeki gözlem değerini, aynı serinin ondan önceki, belli sayıda geçmiş dönem gözlem değerine ve hata terimine bağlı olarak açıklayan modellerdir. AR modeller, içerdikleri geçmiş dönem gözlem değeri sayısına göre isimlendirilirler. Yani; AR modeli bir tane geçmiş dönem gözlem değeri ile ilgileniyorsa "birinci dereceden", iki tane geçmiş dönem gözlem değeri ile ilgileniyorsa "ikinci dereceden" ve genel olarak p tane geçmiş dönem gözlem değeri içeriyorsa "p. dereceden" AR model olarak isimlendirilmektedir.

AR(p) Modeli;

$$x_t = \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \dots + \varphi_p x_{t-p} + e_t \quad (2.1)$$

şeklinde ifade edilir.

Burada $x_{t-1}, x_{t-2}, \dots, x_{t-p}$ gözlem değerlerini, $\varphi_1, \varphi_2, \dots, \varphi_p$ modelin parametrelerini ifade etmektedir. Burada hata terimleri üzerine konmuş varsayım, $N(0, \sigma^2)$ biçimindedir. AR(p) modeli için tahmin edilmesi gerekli parametre ($\mu, \varphi_1, \varphi_2, \dots, \varphi_p, \sigma^2$) sayısı (p+2) tanedir. Uygulamada sıkça kullanılan AR modelleri, birinci ve ikinci dereceden modellerdir. Bunlar AR(1) ve AR(2) biçiminde gösterilir. AR modeller, fark denklemi şeklinde de yazılabilir;

$$Bx_t = x_{t-1}, B^2 x_t = x_{t-2}, \dots, B^p x_t = x_{t-p} \quad (2.2)$$

ifadelerinden yararlanarak, p . dereceden bir AR modeli;

$$x_t = (\varphi_1 B + \varphi_2 B^2 + \dots + \varphi_p B^p)x_t + e_t \quad (2.3)$$

veya,

$$(1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p)x_t = e_t \quad (2.4)$$

şeklinde yazılabilir.

Burada B, geriye öteleme işlemidir (backward shift operator). AR modellerde durağanlık koşulu, polinomun sıfıra eşitlenmesiyle bulunacak köklerin birim çemberin dışında

kalmasıyla mümkündür. Eğer söz konusu kökler birim çemberin dışında kalıyorsa, AR(p) durağan zaman serileri için kullanılabilir. Geriye doğru öteleme işleci kullanılarak AR(1) modeli;

$$(1 - \varphi_1 B)x_t = e_t \quad (2.5)$$

biçimindedir.

Bu modelin durağan olması için,

$$|\varphi_1| < 1$$

koşulunun sağlanması gerekir.

Geriye öteleme işleci kullanılarak AR(2) Modeli;

$$(1 - \varphi_1 B - \varphi_2 B^2)x_t = e_t \quad (2.6)$$

biçiminde ifade edilir.

Bu modelin durağan olması,

$$\begin{aligned} \varphi_2 + \varphi_1 &< 1 \\ \varphi_2 - \varphi_1 &< 1 \\ |\varphi_2| &< 1 \end{aligned} \quad (2.7)$$

eşitsizliklerinin sağlanması gerektirir.

Otoregresif modelin parametrelerinin tahmini, en küçük kareler yöntemi ile yapılmaktadır. Tahmin edicileri elde etmek için p'inci dereceden otoregresif modelin kurulması gerekir. Buna göre ortalaması μ olan p'inci dereceden AR(p) modeli şu şekilde yazılabilir.

$$x_t - \mu = \varphi_1(x_{t-1} - \mu) + \varphi_2(x_{t-2} - \mu) + \dots + \varphi_p(x_{t-p} - \mu) + e_t \quad (2.8)$$

Formülde yer alan x_1, x_2, \dots, x_n gözlemleri; $\mu, \varphi_1, \varphi_2, \dots, \varphi_p$ parametreleri göstermektedir.

$X_t = x_t - \mu$ dönüşümü yaparsak, aynı formül;

$$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \dots + \varphi_p X_{t-p} + e_t \quad (2.9)$$

şekline dönüşür. Böylece parametrelerin en küçük kareler tahminleyicileri aşağıda verilen fonksiyonun minimizasyonu ile elde edilir.

$$S = \sum_{t=p+1}^n [(x_t - \mu) - \varphi_1(x_{t-1} - \mu) - \dots - \varphi_p(x_{t-p} - \mu)]^2 \quad (2.10)$$

veya

$$S = \sum_{t=p+1}^n [X_t - \varphi_1 X_{t-1} - \varphi_2 X_{t-2} - \dots - \varphi_p X_{t-p}]^2 \quad (2.11)$$

AR(1) modelini ele alarak, en küçük kareler tahminlerini yazmak gerekirse model;

$$x_t - \mu = \varphi_1(x_{t-1} - \mu) + e_t \quad (2.12)$$

iken en küçük kareler fonksiyonu,

$$S = \sum_{t=2}^n [x_t - \mu - \varphi_1(x_{t-1} - \mu)]^2 \quad (2.13)$$

olmakta ve bu fonksiyonu minimum yapan en küçük kareler tahminleri aşağıdaki formüllerle hesaplanmaktadır.

$$\mu = \frac{x_{(2)} - \varphi_1 x_{(1)}}{1 - \varphi_1} \quad (2.14)$$

... bunlar neyi ifade ediyor?

$$\varphi_1 = \frac{\sum_{t=1}^{n-1} (x_t - \mu)(x_{t+1} - \mu)}{\sum_{t=1}^{n-1} (x_t - \mu)^2} \quad (2.15)$$

p.'inci dereceden AR modeller için parametre tahminleri de benzer şekilde yapılmaktadır. Parametrelerin bulunmasında diğer bir yöntem de otokovaryans ve otokorelasyon katsayılarının kullanılmasıdır. Bu çalışma içerisinde bu yöntemlere yer verilmeyecektir. Daha geniş bilgi için *Box-Jenkins (1976)*' ya bakılabilir.

... nasıl kullanılır?

2.1.2. MA Modeller

MA modelleri bir zaman serisinin herhangi bir dönemdeki gözlem değerini aynı dönemdeki hata terimi ve ondan önceki belirli sayıda dönemin hata terimine bağlı olarak açıklayan modellerdir. MA modelleri de içerdikleri geçmiş dönem hata terimi sayısına göre isimlendirilmektedir.

q' uncu. dereceden MA(q) modeli;

$$x_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q} \quad (2.16)$$

biçimindedir.

x_t , ($x_t = X_t - \mu$) işlemi ile elde edilen küçültülmüş değerdir, $\theta_1, \theta_2, \dots, \theta_q$ modelin parametreleridir. Bu parametreler, $e_t, e_{t-1}, \dots, e_{t-q}$ değerleri arasındaki ilişkiyi gösteren katsayılarıdır. q ise MA modelinin derecesini gösterir. MA modelinde hesaplanması gerekli parametre ($\mu, \theta_1, \theta_2, \dots, \theta_q, \sigma^2$) sayısı (q+2) tanedir. Uygulamada en çok kullanılan MA modelleri birinci ve ikinci dereceden modellerdir. Bu modeller sırasıyla MA(1) ve MA(2) şeklinde ifade edilir.

MA(1) modelinin yazılımı;

$$x_t = e_t - \theta_1 e_{t-1} \quad (2.17)$$

biçimindedir.

Yazılımdan anlaşılacağı gibi, MA(1) modelinde bir zaman serisinin x_t gözlem değeri t, t-1 dönemlerine ilişkin hata terimlerinin doğrusal bir bileşimidir.

MA(2) modeli;

$$x_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} \quad (2.18)$$

gibidir.

MA modelleri yukarıda gösterilen fark denklemi biçiminde yazılabileceği gibi, "geriye öteleme işleci" B kullanılarak da yazılabilir.

$$x_t = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_p B^p) e_t \quad (2.19)$$

MA modeller için de durağanlık koşulları yerine çevrilebilirlik koşullarından söz edilir. Çünkü MA(1) modeli, sonsuz dereceden AR modeline eşittir. Bunun yanında hata terimine dayanarak bir zaman serisinin herhangi dönemine ilişkin gözlem değeri, sonsuz sayıda hata teriminin tartılı toplamları alınarak açıklanabilir. Ancak uygulama imkansızlığı nedeniyle MA modellerinde zaman serilerinin herhangi bir dönemine ilişkin gözlem değeri, sonlu sayıda hata teriminin tartılı toplamı alınarak açıklanmaya çalışılmaktadır.

MA(1) modeli için "çevrilebilirlik koşulu",

$$|\theta_1| < 1$$

eşitsizliğinin sağlanması ile mümkündür.

MA(2) modeli için bu koşul,

$$\theta_1 + \theta_2 < 1$$

$$\theta_1 - \theta_2 < 1$$

$$|\theta_2| < 1$$

(2.20)

eşitsizlikleri ile ifade edilir.

MA(1) modeli için parametre tahmini;

$$e_t = (1 - B\theta_1)^{-1} x_t \quad (2.21)$$

e_t nin θ_1 parametresine göre kısmi türevi;

$$\frac{de_t}{d\theta_1} = B(1 - B\theta_1)^{-2} \quad (2.22)$$

olur ve bilinmeyen θ parametresinin fonksiyonudur. Bu tür modellerin parametrelerini tahmin etmek için en küçük kareler yöntemi doğrudan uygulanmaz. En küçük kareler yönteminin uygulanabilmesi için modellerin logaritmik dönüşümle doğrusallaştırılması gerekir.

2.1.3. ARMA Modeller

ARMA modelleri durağan zaman serilerinin modellenmesinde kullanılan AR ve MA modellerinin bir kombinasyonudur. Bu nedenle ARMA modellere "karışık modeller" denir. Bu modellerde bir zaman serisinin herhangi bir dönemine ait gözlem değeri, ondan önceki belirli sayıda gözlem değerinin ve hata teriminin doğrusal bir bileşimi olarak ifade edilir. Eğer ARMA modeli; p terimli AR ve q terimli MA modelinin bir kombinasyonu ise, (p+q) terim içerir ve ARMA(p, q) şeklinde yazılır. ARMA(p, q) modeli aşağıdaki gibi ifade edilir;

$$x_t = \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \dots + \varphi_p x_{t-p} + \theta_0 e_t - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q} \quad (2.23)$$

ARMA modeli için tahminlenmesi gerekli parametre ($\mu, \varphi_1, \varphi_2, \dots, \varphi_p, \theta_1, \theta_2, \dots, \theta_q, \sigma^2$) sayısı (p+q+2) tanedir. ARMA modellerinin durağanlık ve çevrilebilirlik koşulunu sağlayıp sağlamadığını belirlemek için bu modelleri "geriye doğru öteleme işleci" B ile yazmak gerekir. ARMA(p,q) modeli B işleci kullanılarak aşağıdaki şekilde yazılabilir.

Bu durumda (2.23) modeli;

$$\varphi(B)x_t = \theta(B)e_t \quad (2.24)$$

biçiminde yazılabilir.

Burada $\varphi(B)$ ile $\theta(B)$ sırasıyla p ve q dereceden polinomlardır. $\varphi(B)$ polinomunun kökleri olan $\varphi_1, \varphi_2, \dots, \varphi_p$ değerleri, birim çemberin dışında kalıyorsa model durağanlık koşulunu, $\theta(B)$ polinomunun kökleri olan $\theta_0, \theta_1, \dots, \theta_q$ değerleri birim çemberin dışında kalıyorsa çevrilebilirlik koşulunu sağladığından söz edilebilir.

2.1.4. ARIMA Modeller

Buraya kadar ele alınmış (B.J) modeller sadece durağan zaman serileri analizinde kullanılan modellerdir. Ancak uygulamada karşılaşılan serilerin çoğu, özellikle ekonomik zaman serileri durağan değildir. Serilerin durağanlığı trend, mevsimsel, konjonktürel dalgalanma ve tesadüfi sebepler gibi etkenler tarafından bozular. Bu etkenlere rağmen zaman serilerinin çoğunda homojenlik görülmektedir (*Box-Jenkins, 1976*).

Durağan olmayan seriyi durağan hale getirmek için uygun derecede fark alma işlemi yapılır. Fark alma derecesi d ile simgelenir ve uygulamada genelde $d=1$, en çok 2 değerini alır. Fark alma derecesi $d=0$ olduğunda bir başka deyişle, seri orijinal değerler itibariyle durağan ise ARIMA model, bir ARMA modele dönüşür.

ARIMA modelinin genel ifadesi;

$$\omega_t = \varphi_1 \omega_{t-1} + \varphi_2 \omega_{t-2} + \dots + \varphi_p \omega_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q} \quad (2.25)$$

biçimindedir.

$\omega_t = \Delta^d x_t$: farkı alınmış seri,

Δ : fark işleci,

d : fark alma derecesidir.

Eğer birinci farklar seriyi durağan hale getiriyorsa fark operatörünün işleyişi;

$$\Delta x_t = \omega_t = x_t - x_{t-1} \quad (2.26)$$

şeklinde gösterilir.

ARIMA(p,d,q) modelinde p veya q sıfır olabilir. Bu durumda model ya AR(d,p) veya MA(d,q) model türüne indirgenmiş olur. Uygulamada sık karşılaşılan $d=1$ veya $d>1$ koşulunu sağlayan bazı ARIMA modelleri fark denklemi şeklinde ve B operatörü kullanılarak aşağıdaki gibi verilmiştir.

i) ARIMA(0,1,1) modeli;

$$\begin{aligned}\Delta x_t &= e_t - \theta_1 e_{t-1} \\ &= (1 - \theta_1 B) e_t\end{aligned}\tag{2.27}$$

eşitlikleri ile yazılır.

Bu modelde $p=0$, $d=1$; $\varphi(B) = 1$; $\theta(B) = 1 - \theta_1 B$ dir.

ii) ARIMA(0,2,2) modeli;

$$\begin{aligned}\Delta^2 x_t &= e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} \\ &= (1 - \theta_1 B - \theta_2 B^2) e_t\end{aligned}\tag{2.28}$$

eşitlikleri ile gösterilir.

Bu modelde $p=0$, $d=2$, $q=2$; $\varphi(B) = 1$ $\theta(B) = 1 - \theta_1 B - \theta_2 B^2$ dir.

iii) ARIMA(1,1,1) modeli;

$$\begin{aligned}\Delta x_t - \Delta x_{t-1} &= e_t - \theta_1 e_{t-1} \\ (1 - \theta_1 B) \Delta x_t &= (1 - \theta_1 B) e_t\end{aligned}\tag{2.29}$$

eşitlikleri ile ifade edilir.

Bu modelde $p=1$, $d=1$, $q=1$; $\varphi(B) = 1 - \varphi_1 B$; $\theta(B) = 1 - \theta_1 B$ dir.

ARIMA modellerde durağanlık ve çevrilebilirlik koşulunun sağlanıp sağlanmadığını belirleme işlemi ARMA modellerde olduğu gibidir. Gözlem değerlerinden oluşan zaman serilerinin birbirini izleyen yılların aynı aylarında veya dönemlerinde maksimuma ve minimuma ulaşma eğilimi mevsim dalgalanmalarını ifade etmektedir.

Doğal ve sosyal nedenler sonucu ortaya çıkan ve her yıl düzenli olarak tekrar eden bu dalgalanmaları içeren serilere "mevsimsel zaman serileri" adı verilir. Aylık gözlem değerlerinden meydana gelen serilerde genellikle, mevsimsel dalga uzunluğu $s=12$ 'dir. Ancak 6 aylık periyoda sahip mevsimsel dalgalanmalara da rastlanabilir. Üç'er aylık aralıklarla yapılan gözlem değerlerinden oluşan serilerde $s=4$ ' tür.

2.2. Modelin Uygunluk Testi

Seçilen modelin ele alınan seri için uygun olup olmadığının araştırılması gerekir. Minimum parametrelili uygun modelin seçilmesinde kullanılan testlerde, tahmin hatalarının otokorelasyonlarına dayalı testler daha pratiktir. Hata otokorelasyonlarının her birini ayrı ayrı kendi standart hataları ile $1/\sqrt{n}$ değerini karşılaştırmak mümkündür. Ancak tahmin edilen otokorelasyonların standart hatasını bir ölçü olarak kullanmak, düşük derecelerdeki gecikmelerde hesaplanan otokorelasyonların sıfırdan farklılığının önemini açıkça ortaya koyamaz (Kayım, 1985). Bu nedenle otokorelasyonları tek tek incelemek yerine belirli sayıdaki otokorelasyonları bir arada incelemek daha başarılı sonuçlar vermektedir. Bu şekilde yapılan inceleme ile seçilen modelin uygunluğunu daha açıkça ortaya çıkarabilmektedir. Bu amaçla *Box-Pierce* tarafından geliştirilmiş olan *Box-Pierce* istatistiği Q şu şekilde hesaplanmaktadır (Box-Jenkins, 1976).

$$Q = \sum_{k=1}^K r_k^2(e) \quad k = 1, 2, \dots, K \quad (2.30)$$

Formülde yer alan $r_k(e)$, örnekleme içi tahmin hatalarının çeşitli k gecikmelerdeki otokorelasyonlarıdır. Burada n, örnekleme hacminden fark derecesinin çıkarılması sonucu kalan değerdir.

Box-Pierce istatistiği Q, yaklaşık olarak Ki-kare dağılımlı olup, (K-p-q) serbestlik derecesine sahiptir. K, hesaplanan otokorelasyon sayısı p, otoregresif q, hareketli ortalama modelindeki parametre sayılarıdır. Bu testlerde ki-kare dağılımı kullanıldığından uygulamada ki-kare testi olarak bilinir. Bu test, bazı otokorelasyonların anlamlı olarak sıfırdan farklı olup olmadıklarını ortaya koyar. Bu test ile "otokorelasyon ortalaması sıfırdan büyüktür" hipotezi test edilir. Bu hipotezin reddi, hataların rastgele dağıldığını, uygulanan yöntemin ve seçilen modelin uygun olduğunu ifade eder.

2.3. Box-Jenkins Yönteminin Üstün ve Zayıf Yönleri

2.3.1. Üstün Yönleri

i) (B.J) yönteminde uygun modelin belirlenmesi genellikle eldeki verilerin yapısı ile belirlendiği için "verilerin kendi kendine ilişkisi sağlanmış olur". Bu nedenle (B.J) modellerine dayanarak yapılan kısa dönem tahminlerinin diğer yöntemlere dayanarak yapılan aynı döneme ait tahminlere oranla daha güvenilir olduğu söylenebilir.

ii) (B.J) yönteminde ileriye dönük tahmin amacıyla analiz edilecek bir zaman serisi için uygun model belirlenirken izlenen her aşamada bu modelin analiz edilecek seriye uygunluğunu denetleme imkanı vardır.

iii) (B.J) yöntemine göre belirlenecek uygun modelde önemli olan parametre sayısının olabildiği kadar az tutmaktır.

iv) Zaman serilerinin çoğunda ardışık gözlem değerleri birbirine bağımlıdır. (B.J) yöntemi zaman serilerinin bu en önemli özelliğini en etkili biçimde kullanır(Özmen, 1986).

2.3.2. Zayıf Yönleri

i) (B.J) yöntemine dayanarak yapılan tahminler çabuk elde edilemez. Çünkü (B.J) yöntemi tümüyle otomatik değildir. Bu yöntemle dayanarak ileriye dönük tahmin yapmak amacıyla yazılacak bir bilgisayar programı yinelemeli bir programdır.

ii) (B.J) yönteminin uygulanabilmesi uzman ve deneyimli işgücüne gereksinim duyulur.

iii) (B.J) yönteminin uygun model seçimi konusunda sağladığı özgürlük olanağı, tahmin yapan kişinin uygun olmayan model seçmesine neden olabilir.

iv) (B.J) modelleriyle aynı seriyi analiz eden ve aynı seri için ileriye dönük tahmin yapan iki kişinin sayısal olarak birbirine benzer sonuçlar elde etmesi konusunda garanti yoktur.

v) Model belirlemek için çok sayıda gözlem değerine gereksinim vardır, bu sayı en az 50 olmalıdır. Box-Jenkins yöntemlerinin bu olumsuzluklarına rağmen, bir zaman serisinin yapısını belirlediği, gözlem değerlerinin aralarındaki bağımlılığı en etkili bir şekilde kullandığı ve model belirleme aşamalarında istatistiksel testlere yer verdiği için, diğer tahmin yöntemlerine göre kısa dönem tahmin yapmada üstün bir yöntemdir (Özmen, 1986).

2.4. Zaman Serilerinin Analizi

Box-Jenkins grubu modeller, durağan karakterde olup olmama durumlarına göre iki grupta incelenir. Durağan özelliğe sahip modeller, ARMA grubu modeller olarak bilinir ve AR(1), MA(1), AR(2), MA(2) ve ARMA(1,1) modelleri bu tip özelliğe sahip modeller olarak tanımlanır. Durağan olmayan modeller grubunda ise, IAR(1,1,0), IAR(2,1,0), IAR(2,1,0), IAR(2,2,0), IMA(0,1,1), IMA(0,1,2), IMA(0,2,2) veya ARIMA(1,1,1) modeller bulunur.

Zaman serileri için uygun modelin seçimi, *Box-Jenkins (1976)* tarafından aşağıda tanımlanan adımlarla gerçekleştirilir.

- Gözlem değerlerini oluşturan seri incelenerek, model grubu kararlaştırılır. Bu aşamada hangi model grubunun uygun olacağı kararlaştırılır : Uygun modeli belirleme aşamasında yapılan ilk iş; seri durağanlığının belirlenmesidir. Durağanlığın irdelenmesinde kullanılan araç otokorelasyon ve bu fonksiyonların korelogramdır.
- Kararlaştırılan model grubundan hangi model tipinin ilgili seri için uygun olacağı belirlenir : Durağan zaman serilerinin analizi ve tahmini aşamalarında üç tür modelden (AR, MA, ARMA) biri kullanılır. Model seçiminde izlenen yol; otokorelasyon ve kısmi otokorelasyon fonksiyonlarının kullanımı ile yapılır. Bu şekilde seçilen model için parametrenin derecesi de belirlenir.

- Uygun olacağına karar verilen modelin parametreleri tahmin edilir : Parametrelerin tahminlenmesi aşaması son derece karmaşık ve zaman alıcı işlemlerin yapılmasını gerektirir. Bu işlemler her model türü için farklılıklar göstermekte ve istatistik paket programlarının kullanımı ile yapılmaktadır.
- Modelin yeterliliği, (2.2) kısmında açıklanan Box-Pierce istatistiği ile test edilir. Model yeterli ise tahmin amacıyla kullanılır. Aksi durumda birinci adıma dönülür.



ÜÇÜNCÜ BÖLÜM

3. SAPAN DEĞERLER

3.1. Temel Tanımlar

Bazı arařtırmalarda bir kitleden alınan örneklem içinde bir ya da birden çok deęer beklenenin aksine ana kitleden farklı olabilir. Farklı gibi görünen gözlem deęerlerinin kitleden çıkarılması veya çıkarılmaması deneyi yapan kiřiye baęlıdır. Eldeki örneklem deęerleri içinden bir çift sapan olmayan deęerin çıkarılması ile meydana gelecek kayıp, bir çift sapan deęerin örneklem seti içerisinde tutulması ile sebep olunan kayıp ile karşılaştırıldığında daha küçük olabilir. Bu durumda tartışılmaz tek nokta, sapan deęer probleminin varlığıdır. Bazı istatistikçiler, "Sapan deęerlerin sapan olma olasılıkları gözlem hakkında bilgi vermiyorsa bu tip gözlemlerle çalışılmamalıdır" demektedir. Bazıları da "Eđer varlığında fiziksel bir açıklama bulunmuyorsa, sapan gözlem kavramı kabul dahi edilmemelidir" düşüncesini savunmaktadır. Ancak, esas olan deneyle uğrařan bilim adamları ve arařtırmacıların sapan gözlemler hakkında "elde tutmak ya da tutmamak, reddetmek ya da reddetmemek" şeklinde karar vermeye zorlanmalarındır. Mesela *Calvin* (1949) tarafından yazılan bir kimya ders kitabı, okuyucularına *Chauvenet*'in ařaęıda tanımlanan yöntemini kullanmalarını tavsiye eder. Bu oldukça ilginç yöntem kısaca şöyledir; "n adet gözlemden oluşan bir dizinin herhangi bir sonucu, tüm ölçümlerin ortalamasından sapmanın büyüklüęü, yeterli büyüklükteki tüm sapmaların ortaya çıkma olasılıęından küçük olduğunda red edilecektir". Bu yöntem, sapan gözlem konusunda varolan en eski kural olarak bilinmektedir (*Collett ve Lewis, 1976*).

Uygulamalı durumlarda genellikle parametrelerin deęerleri, olasılık modellerinin mümkün olan farklı kümeleri ile gösterilir. Örneklemdaki veriler, eldeki modelin geçerlilik derecesini belirlemek veya ilgili parametrelere ait olan hipotezleri test etmek ya da parametreleri tahmin etmek için incelenebilirler. Modeldeki yanlışlık ve istatistiksel yöntemlerin hatalı kullanımı, veri kümesinin doğruluk derecesini belirlemedeki önemini biraz azaltabilir. Bununla birlikte, parametre tahmini için veri süreci ve modelin uygulanabilir olması varsayımının test edilmesi konularına daha çok eğilim

gösterilmektedir. Bu, biraz tehlikeli bir kuraldır. Örneğin, normal dağılım ortalamasını tahmin etmek için hangi yöntemin iyi olduğu, dağılım normal değilse yetersiz olabilir. Veriler normal dağılıştan elde edilen bir örneklem gibi düşünüldüğünde bir ya da daha fazla gözlem şüpheli görünebilir.

Bu evrede "sapan gözlem" ile ne demek istendiği açık hale getirilmelidir. Sapan gözlem, "Bir veri kümesinde, kümenin geri kalanıyla tutarsız gibi görünen bir ya da birden çok gözlemin alt kümesidir". Tutarsız gibi görünen cümlesi çok önemlidir. Dikkatli gözden geçirmek için bazı gözlemleri seçip seçmemek, gözlemci açısından subjektif bir düşünce sorunu olabilir. Gözlemci için endişeye sebep olan husus, bazı gözlemlerin ana kitlenin gerçek elemanları olup olmadığıdır. Eğer gözlemler kitlenin gerçek elemanları değilse, gözlemcinin söz konusu kitle hakkında yorum getirmek için yaptığı çalışmaların anlamı yoktur. Ancak sapan değeri nitelendiren, gözlemci üzerindeki etkisidir. Böyle gözlemler ana kitleden hariç olmalı mı sorusuna cevap bulmak gerekir. Aksi durumda kitleyi oluşturan gözlemlerin kitleyi ne kadar temsil ettiği bilinemez. Bu güçlükten, modeldeki parametre tahminleri olumsuz etkilenebilir.

Sapan gözlem, bir latinkare deney düzeninde gözlenen bağımlı değişkenlerin sayısı ya da değeri olabilir. Modeldeki parametre uyumu sağlandığında ve uyumu sağlanan değerlerden gözlenen bağımlı değişkenlerin sapmaları tablollaştırıldığında gerçek sapmalar hakkında yorum yapılabilir. "Hataların testi, regresyon analizinde üzerinde önemle durulan bir ayrıntı olmalıdır" şeklinde tavsiye edilen hüküm, geniş bir araştırmacı kitlesi tarafından kabul edilmektedir. Bu durum, sapan gözlemleri sapan hata değerleri olarak ele alınması sonucunu doğurmaktadır.

3.2. Kuralsız Yaklaşımlar

Kural dışı ya da şüpheli değerler probleminin varlığı 18'inci yüzyılın ortalarından beri bilinmektedir. *Daniel Bernoulli (1777)*, astronomide biribiri ile ilişkili gözlemler hakkında yazılar yazmıştır. Bundan daha eski bir çalışma *Boscovich (1755)* tarafından yapılmıştır : Söz konusu çalışmada dünyanın elips şeklinde oluşunu saptamak için teleskop üzerinde kutuba ait ısının fazlalığının 10 örneklem ölçümü için, iki uç değeri sapan gözlem gibi analiz dışı bırakarak 8 örnekten ortalamayı hesaplamış ve karar vermiştir. Bu dönemde 19'uncu yüzyılın ortasına kadar sapan gözlemler konusunda tartışılan tek nokta, sapan olan gözlemin seriden atılmasının (reddetme) doğru bir karar olup olmadığıdır.

Bessel ve Baeuer (1888), hatanın büyüklüğünden dolayı bir gözlemi reddetmediklerini, eldeki eşit ağırlıklı tüm verilerin sonuca katkıda bulunmalarını sağladıklarını söylemişlerdir. *Boscovich* ve bir kısım araştırmacı, sapan olan gözlemi seriden çıkarma anlamına gelen reddetmeyi uygulamışlardır. Bununla birlikte o yıllarda reddetme işlemi planlanmış değil, tamamen gözlemcinin kararıyla ilgili bir tasarruf olmuştur. Mesela *Legendre (1805)*, kabul edilebilir olmasına karar verilmiş sapmaların reddedilmelerini tavsiye etmiştir.

Kural dışı gözlemler için yayınlanan ilk objektif test, Amerikalı astronom *Pierce'e (1852)* ait olanıdır. *Pierce* yönteminde, "n gözlemden oluşan bir örnekleme k şüpheli gözlem reddedilmelidir" demektedir. Bu yöntemi, Amerikalı astronom *Chauvenet'in (1863)* tek kuşkulu gözlem değerleri için geliştirdiği yöntem izlemiştir. *Chauvenet'ten* sonra *Stone (1868)* bir test ortaya koymuştur. Bu testin dayandığı fikir "Dikkatsizliğin modülü" biçiminde açıklanmıştır. Bir gözlemci, seçilen bir örnekleme ele aldığı her m gözlemden ortalama olarak bir hata yapar. Eğer gözlemden sapma, gözlemcinin dikkatsizliğine göre rasgele değişimden daha büyük olasılıkla tayin edilebilirse bu gözlem atılabilir.

Uç değerlerin sınırsız olarak reddedilmesi konusunda alternatif testler bir kaç yazar tarafından ele alınmıştır. *Stone*'un reddetme testinden bir kaç yıl sonra bir kaç yöntem daha yayınlanmıştır. Bu testlerden biri *Stone* tarafından ele alınmıştır. Bu test'te örneklem ortalamasını hesaplamada gözlemlerin ağırlıkları esas alınmıştır. Bununla sapan gözlemlerin uyumlarını sağlamlaştıran bir konum parametresinin tahmini için geliştirilen yöntem kullanılmaktadır. Anılan sapan değer testlerine bakıldığında *Wright* (1884) yöntemi dikkate değerdir. Bu test yöntemine göre "Herhangi bir gözlemin ortalamadan sapmasının, standart sapmanın 3 katından büyük olduğunda gözlem reddedilir". Bu test'in biraz değiştirilmiş uyarlaması ile *Wright* ve *Hayford* (1906) tarafından geliştirilen teste göre, "Muhtemel hatayı 3.5 kat aşan hata değeri için her gözlem test edilecek ve eğer birarada bulunan koşullardan herhangi biri, güven yetersizliği oluşturur ise reddetme kuralı uygulanacaktır".

Goodwin (1913) tarafından geliştirilen yöntem ise, "n gözlemden oluşan bir örneklemdeki sapan gözlemin sapma ortalaması, geri kalan n-1 gözlemin ortalamasından sapma ortalamalarını 4 kat aşarsa reddedilir" şeklindedir. *Wright* ve *Goodwin*'in testlerindeki kritik oranlar için örneklem büyüklüğü n'den bağımsız olan değerleri (3.50, 4.00, 5.00) seçtikleri söylenmektedir. Bu durum, şimdiye kadar önerilen bütün test yöntemlerinin genel bir eksikliği ile ilişkilidir. Kuralsız yaklaşımlarda sapan değer testlerinin amacı hakkında verilmeye çalışılan fikir; sapan gözlemlerin temel bir olasılık modelinde olumsuz etki yansıtılabileceği veya yaygın istatistiksel yöntemlerle elde edilmeye çalışılan, yöntemlerin optimal özellikleri ile ilgilidir. Sapan gözlemler ile ilgili teori ve uygulamayı içeren ayrıntılı testler, bu temeller üzerine inşa edilmelidir. Burada sorulması gerekli en önemli soru şudur; "Sapan gözlemlerle neden ilgilenilmektedir?". Sapan gözlemlerle ne demek istendiği hatırlanacak olursa; bu kavramın, araştırmacılar tarafından iki farklı anlamda kullanıldığı görülecektir. Burada kullanıldığı anlamı ile verilerin yerini subjektif olarak ortaya koymasındır. *Grubbs* (1969), "Sapan bir gözlem ortaya çıktığında örneklemdeki diğer gözlemlerden ciddi derecede farklı olan değerdir" demektedir. Bu tip sapan gözlemlerin bulunduğu veri kümeleri için elde edilen parametre değerleri ile, ideal modeli bağdaştırmak mümkün olmaz. Hem elde edilen verilerle ilgili hem de verilerin ortaya çıkışını anlatan uygun bir olasılık modeline göre sapan

gözlemlerin uygunluğu hakkında bazı şüpheler olabilir. Buna göre veriler için üretilen düşüncelerin muhtemel olasılık modelleri ile ne kadar farklı olduğuna dikkat edilmelidir. Normal dağılım beklendiği halde, ilgi uyandırmayacak kadar kuşkulu olan gözlemlere tepki gösterilmeyebilir. Sapan gözlemleri test etmek için kullanılan istatistiksel testin amacı, belirli bir veri kümesinde bulunan sapan gözlemlerin varlığının subjektif düşünce yerine, veri analizi için önemli objektif yorumlara sahip olunup olunmadığına karar vermeyi sağlamaktadır.

Sapan gözlemler sadece belirleyici sebeplerden dolayı ortaya çıkmış olabilirler. Bu belirleyici sebepler, okuma yazma ya da hesap hatası olabilir. Eğer durum böyle ise, çözüm basittir. Yanlış yapılmış örneklem verileri çıkarılmalı veya düzeltme yöntemi ile düzeltilmiş veriler örnekleme dahil edilmelidir. Daha kesin olmayan durumlar için sapan gözlemler hakkında uyumsuzluğu giderecek uygun istatistiksel dönüşümler kullanılmalıdır.

Bazı araştırmacılar, "sapan gözlem" kelimesini hem hayret verici hem de uyumsuz gözlemler için kullanırlar. *Grubbs (1950)* 'nin ifadesi tipik bir örnek olarak şöyledir; "en büyük ve en küçük gözlemler örnekleme diğer gözlemlerden farklı görülebilir". Burada hem en büyük hem en küçük gözlemlerin gerçekten sapan gözlemler olduğu hipotezinin test edilmesi ile ilgilenilmektedir. Bir veri kümesinde uç değerler her zaman bulunabilir, önemli olan bu değerlerin sapan olup olmadıklarıdır. Bu değerler uygun görülen modelden tesadüfen elde edilmiş olmayabilir. Gözlem setindeki söz konusu hata için, tahmini gerçekleştiren sebepler bulunabilir ve yanlış düzeltilebilir. Uyumsuzluğu tespit etmede kullanılan istatistiksel yöntemler, sapan gözlemlerle ilgili kaynaklara büyük katkı yapmıştır. Kaçınılmaz olarak sapan gözlemler için uyumsuzluk testi, veri analizinde önemli bir rol oynar. Verilerde uyumsuzluk problemini ortadan kaldırmak için karmaşık istatistiksel yöntemler yerine basit yöntemlerle gözlemlerdeki uyumsuzluğu test etmek daha etkili olabilir. Bu tip analizler, sapan gözlem araştırmasının ilk evresinde yapılmalıdır. Bir ya da daha fazla gözlemin uyumsuzluğuna karar verilirse, bu veriler için uygun tahminler yapılabilir veya gözlemler seriden çıkarılabilir. Bir başka durum olarak, orijinal modelde değiştirilmiş "hata" verileri için analizi geliştirmeye karar verilebilir veya

uygun sapan gözlemleri birarada toplamak için model değiştirilebilir. Ya da, sapan gözlemleri uyumlu hale getirmek için tüm veri kümesini analiz eden bir yola başvurulabilir.

Sapan gözlemler konusunda eleştiri yazıları yazan bazı yazarlar sapan gözlemlerin ortaya çıkabileceği durumlar üzerine tartışmışlardır. Bu durumlar üç grupta ele alınmıştır.

- **Doğal Değişim**

Gözlemlerin kitle üzerinde doğal değişiminden kaynaklanmaktadır. Bu tip değişimler kontrol edilemezler. Örneğin; insanlara ilişkin boy ölçümleri, göz renkleri, zekaları ve benzeri özelliklerle değişimlerin ifade edildiği durumlar.

- **Ölçüm Hatası**

Ölçüm aletinde meydana gelen sapmalar, ayrıca verilerin yuvarlanması veya kaydetme hatası gibi kişiden kaynaklanan hataların sebep olduğu durumlar doğasında varolan etken üzerindeki değişimin bir başka boyutunu ortaya koyar. Bu gibi değişimlerin kontrolü genelde mümkündür.

- **Uygulama Hatası**

Değişimin başka bir kaynağı olan uygulama hatası, verilerin eksik toplanmasından ortaya çıkar. Yanlı örneklem seçimi veya kitleyi temsil etmeyen bir verinin dahil edilmesi durumunda bu tip hatalar oluşabilir. Alınacak önlemlerle kısmen de olsa bu tip hataları ortadan kaldırmak mümkündür.

Değişimin bu üç tipiyle ilişkili sapan gözlemler sınıflandırılabilir. Bir veri kümesindeki sapan gözlem, aslında doğasında varolan değişimin tamamen mantıklı bir yansıması olabilir. Eğer istatistiksel olarak uyumsuz bir durum ortaya çıkar ise, bu durum uygun olmayan model durumunu ifade eder. Burada üzerinde önemle durulması gerekli husus, bariz bir ölçüm hatası, gözlemin reddini destekler.

Uygulama hatası olarak ortaya çıkan bir sapan gözlem bazen reddetmeyle sonuçlanabilir. Doğal rasgelelik veya ölçüm tartım ve kaydetme hatasından ortaya çıkan gözlemler arasındaki farklılık terimleri teknik olarak ayırmıştır. Sapan gözlemler için yapılan bütün istatistiksel yöntemlerde ortaya çıkan değişim, ayrıntılı bir biçimde ifade edilmelidir. Ancak, kişilerden kaynaklanan hatalardan ortaya çıkan gözlemler kabul edilmez. Bu durumda reddetme yada düzeltme işlemi yapılması kaçınılmazdır. Çalışmaların bir çoğunda karşılaşıldığı gibi, veri dizisindeki bazı değerler diğerlerinden oldukça büyük farklılıklar göstermelerine karşın bunların sapan değerler olduğunu ancak bazı yaklaşık testlerle söylemek mümkündür. Eğer sapan gözlemler doğal rastgelelikler sonucu ortaya çıkmış ise, bu durumda gözlemler atılmaz, bunlara da örneklemdaki diğer gözlemlere uygulanan işlemler uygulanır. Eğer sapan değerlerin oluşmasına; önceden düşünülen deneysel yöntemden büyük miktarda sapma, hesaplamadaki yanlışlıklar veya sayısal değerlerin yanlış kaydedilmesi gibi kişisel hatalar, ölçüm aletlerinin değişimi veya ayarlarının bozulması, ölçüm zamanlarının değişikliği neden olmuş ise bu durumda sapan değerlerin örneklemden çıkartılması, yerine yenilerinin çekilmesi, gerektiğinde ortalama yerine ortancanın kullanılması geçici de olsa birer çözümdür.

Sapan değerlerin bulunmasında kullanılan testlerin en belirgin olanları ana başlıklarla aşağıdaki gibidir;

- Varyansı bilinen normal dağılımlı örneklemlerde bir veya iki sapan değer testi:
Burada, varyansı bilinen normal dağılımlı örneklemlerin sapan değerlerini incelemek amacıyla iki işlem yapılmaktadır.
 - i) Gözlemlerden hata değeri maksimum olan gözlem için önceden belirlenen sabit aşıyorsa, sözkonusu gözlem değerine sapan değer denir ve aynı işlemler sıra ile diğer gözlem değerlerine uygulanır.
 - ii) En büyük veya en küçük iki gözlemin hata değerleri toplamı önceden belirlenen bir sabiti aşıyorsa bu gözlemlere sapan değer denmektedir.

-
- Varyansı bilinmeyen normal dağılımlı örneklerde bir veya iki sapan değer için yapılan testler:

Bu testler için aşağıda tanımlanan üç farklı yol izlenir;

i) En büyük gözlem değerinin sapan olabilmesi için, artık değerlerin kendi tahmini standart sapmalarına bölünmesi ile elde edilen standartize (studentize) artık değerden en büyüğünün, önceden belirlenen bir değeri aşması gerekmektedir. Bu şekilde bulunan sapan değer örneklerden çıkartılır ve aynı işlemler gözlemlerdeki sapan değerlerin tümü için tekrarlanır.

ii) En büyük veya en küçük iki gözlemin sapan gözlem olabilmesi için, en büyük iki standartize artık değer toplamının önceden belirlenmiş bir değeri aşması gerekmektedir.

iii) İki büyük gözlem değerinin sapan olabilmesi için bu değerlerin ihmal edilmesi durumunda bulunan düzeltilmiş kareler toplamının, kareler toplamına oranının bir kritik orandan küçük olması gerekmektedir.

- Doğrusal modellerdeki sapan değerlere ilişkin testler:

Bu tür testlerin en önemlisi, standartize değerlerin ve artık değerlerin sabit standart sapmayla standartlaştırılmasından bulunan bir test istatistiğinin kritik değerlerine ilişkindir. Bu test her türlü doğrusal modele uygulanabilmektedir.

Doğrusal modellerdeki sapan değerlerin yaklaşık testlerinde kullanılan çeşitli tablolar geliştirilmiştir. Bir örnekteki en büyük artık değer, bir sapan değer olup olmadığı 0.10, 0.05 ve 0.01 seviyelerindeki kritik tablo değerleri ile karşılaştırılarak yapılmaktadır.

İstatistiksel analizlerde parametre kestirimi yapılmadan önce, gözlem değerleri arasında sapan değerlerin bulunup, bulunmadığının araştırılması gerekir. Çünkü bu tür farklı değerlerin seri içerisinde bulunması durumunda elde edilecek parametre değerleri farklı ve buna bağlı olarak elde edilen ileriye doğru kestirimler gerçeği yansıtmaktan uzak olmaktadır.

Tarihi gelişimi yukarıda belirtilen basit örneklemeler için sapan değer analizi çalışmaları daha sonraları deney düzenleme, çok değişkenli istatistik analiz, regresyon analizi ve zaman serileri modellerini kapsayacak şekilde genişlemiş ve istatistiği kullanan bütün bilimlerde göz önüne alınmaya başlanmıştır. Yapılan bilimsel çalışmalarda sapan değerlerin gözönüne alınmaması yanlış parametre kestirimleri ve geniş güven aralıkları elde edilmesine sebep olmuştur. Bu nedenle sapan değer analizi, bilimsel bütün çalışmalarda yapılacak istatistik analizlerin sağlıklı yürümesi bakımından temel analiz özelliği taşır.

Sapan değer analizi çalışmalarında, eldeki seri değerleri arasından sapan olanları bulurken testin gücü çok önemlidir. En başarılı ve güçlü testlerin bile sapan değer sayısı çok olan bir seri için etkili olmadığı, sapan değer sayısının artmasına paralel, tarama süreçlerinin gücünde zayıflama sonucunun ortaya çıktığı tespit edilmiştir. Bunun yanında seri içerisinde tek sapan değer olması durumunda, zayıf bir test yönteminin bile başarılı sonuçlar verdiği belirlenmiştir. Bu çıkarımlar simülasyon çalışmalarıyla *Chang, Tiao ve Chen (1988)* tarafından da ortaya konmuş ve doğrulanmıştır.

3.3. Basit Örneklem için Sapan Değer Analizi

Basit örneklem için sapan değer testinde kullanılan yöntemlerden bazıları bir yada iki sapan gözlem değerinin testinde kullanılırken, bazıları çok sayıda sapan gözlem değerinin testine olanak sağlamaktadır. Bunlardan yaygın olarak kullanılan yöntemler;

- Thompson Yöntemi,
- Pearson ve Chandra Sekar Yöntemi,
- Grubbs-I Yöntemi,
- Dixon Yöntemi,
- Ferguson Yöntemi,
- Grubbs Yöntemi-II,
- Rosner Yöntemi ve diğerleri şeklinde sıralanabilir. Daha geniş bilgi almak için *Tatlıdil (1981)*'e bakılabilir.

• Thompson Yöntemi

Thompson tarafından 1935 yılında önerilen yöntem bu konudaki ilk matematiksel yöntemlerden biridir. Yöntemde x_1, x_2, \dots, x_n gibi normal dağılımlı kitleden çekilmiş bir örneklem için ortalama \bar{x} , varyans s^2 , herhangi bir gözlem değeri x_i olmak üzere,

$$T_i = \frac{(x_i - \bar{x})}{s} \quad i = 1, 2, \dots, n \text{ için} \quad (2.31)$$

test istatistiği önerilmiştir. Bu istatistiğin dağılımı araştırılarak çeşitli n ve α değerleri için

$$P(|x_i - \bar{x}|/s > \tau_\alpha) = \alpha \quad i = 1, 2, \dots, n \quad (2.32)$$

eşitliğinden τ_α ölçüt değerleri simülasyon yöntemi kullanılarak bulunmuştur. Bu koşullarda $|x_i - \bar{x}|/s > \tau_\alpha$ sonucunu veren x_i gözlem değerinin "sapan değer" olduğuna karar verilmektedir.

- **Pearson ve Chandra Sekar Yöntemi**

Pearson ve Chandra Sekar'ın 1936'da geliştirdikleri yöntem, örneklemedeki kuşku edilen bir en büyük ve bir en küçük gözlem değerinin sapan olup olmadıklarının testinde kullanılmaktadır. Örneklemedeki değerler $x_{(1)} < x_{(2)} < \dots < x_{(n)}$ biçiminde küçükten büyüğe sıralanmaktadır. Kuşku edilen gözlem değeri $x_{(n)}$ ise, (2.31)'de ifade edilen eşitlik kullanılarak test işlemi yapılır.

- **Grubbs-I Yöntemi**

Pearson ve Chandra Sekar tarafından 1936'da önerilen yöntemi, Grubbs, 1950 yılında geliştirerek, bir en büyük veya bir en küçük gözlem değerinin sapan değer olup olmadığını testinde kullanılmak üzere önermiştir.

- **Dixon Yöntemi**

Gözlem sayısının büyük olduğu örneklerde standart sapma ya da varyansın hesabı işleminin güç olduğu düşüncesiyle Dixon, gözlemler arası uzaklığın örnekleme genişliğine bölümünden bulunan bir dizi test istatistiği önermiştir.

- **Ferguson Yöntemi**

Ferguson tarafından 1961'de önerilen yöntemler tek yanlı testlerde % 50'den az bozulmalar için geçerli ve gizlenme etkisinden etkilenmeyen çarpıklık katsayısı ile ve yine gizlenme etkisinden etkilenmeyen iki yanlı testlerde % 21'e kadar bozulmalar için geçerli, basıklık katsayısı üzerine kuruludur. Ferguson ayrıca Monte-Carlo çalışması ile tablolar geliştirmiştir.

- **Grubbs-II Yöntemi**

Çok sayıda kuşku gözlem değerlerinin aynı anda testine olanak sağlayan ilk yöntem Grubbs tarafından 1969'da geliştirilmiştir ve bunlar Grubbs-Türü test istatistikleri olarak bilinir. Normal dağılımlı kitleden çekilmiş n denekli bir örneklemedeki k adet kuşku gözlem değerinin sapan değer olup olmadıklarının testinde kullanılan bu yöntemlerden ilkinde gözlem değerleri $x_1 < x_2 < \dots < x_n$ biçiminde sıralandıktan sonra en

büyük k tane gözlem değeri için test işlemi yapmaktadır. Bu yöntem kuşkulu k tane gözlem değeri örneklemin bir ucunda ise geçerlidir. Kuşkulu k tane gözlem değerinin örneklemin her iki ucunda da olması durumunda ise benzer bir yöntem önerilmektedir. Grubbs, önerdiği test istatistiklerinin üst sınırlarını çeşitli α ve n değerleri için benzetim teknikleri kullanılarak tabloştırmıştır.

- **Rosner Yöntemi**

Rosner'in 1975 ve 1977 yıllarındaki çalışmalarında önerdiği R istatistiği yöntemi, normal dağılımlı kitlelerden çekilmiş örneklemden çok sayıda kuşkulu gözlem değerinin testinde kullanılan en güvenilir yöntemlerden birisidir.



3.4. Zaman Serilerinde Sapan Değerler Analizi

Çeşitli veri analizlerinde bazı gözlemlerin diğerlerine uyum göstermediği durumlara çok sık rastlanır. Grevler, savaşlar, bir malın piyasasında veya fiziksel sistemde meydana gelebilecek beklenmeyen değişiklikler ve benzer durumlardan ötürü bazı gözlemler sapan olabilir. Belirtilen durumlar, beklenmeyen müdahale olarak tanımlanmaktadır. *Gutman ve Tiao (1978)*, *Miller (1980)* ve *Chang (1980)* beklenmeyen müdahalelerin; otokorelasyonlar, kısmi otokorelasyonlar ve ARMA parametrelerini tahminlemede ciddi şekilde yanlılığa sebep olduklarını göstermişlerdir. Dışsal etkileri tanımlayabilmek ve gözlemlerde bu etkileri giderebilmek için seri yapısının çok iyi bilinmesi gerekmektedir. Eğer müdahale anı biliniyorsa, *Box ve Tiao (1975)* tarafından geliştirilen ve bu bölümün (3.4.1) kısmında tanımlanan "Müdahale Analiz Yöntemi" kullanılarak, sözkonusu etkinin büyüklüğü belirlenebilir (*Chang, Tiao ve Chen, 1988*).

Sapan değer tarama yöntemlerinden bazıları, özellikle basit örneklem için geliştirilenleri, tek sapan değer tarama esasına göre çalışan, sapan değerleri belirleme yönünden çok güçlü olmayan yöntemlerdir. Bu yöntemler, her adımda tek sapan değer tespit esasına dayanır ve "adımsal test yöntemleri" olarak bilinirler. Oysa daha etkin olan yöntemler, olması muhtemel bütün sapan değerlerin tespit edilmesine yöneliktir ve bunlar "aynı anda test yöntemleri" olarak tanımlanmaktadır. Yukarıda karşılaştırılan test yöntemleri de bu gruba girmektedir. Her iki grup test yönteminin de bazı sakıncaları vardır. Adımsal test yöntemleri olarak bilinen birinci grup testlerin uygulamasında, eğer örnekleme bulunan sapan değerler birbirlerine pozisyon olarak yakın iseler, uzaklık (distance) ölçütlerine dayalı yöntemler kullanıldığında (regresyon analizinde sapan değerler için Cook yaklaşımı gibi) gerçekte sapan olan bazı gözlemlerin "gizlenme etkisiyle" (masking effect) sapan olmadığı sonucuna varılabilir. Aynı anda test yöntemleri gizlenme etkisinin giderilmesine yönelik olarak geliştirilmelerine karşın bazı olumsuzluklar taşımaktalar. Bu yöntemlerde, örnekleme bulunan sapan değerler çok farklı iseler, gerçekte sapan olmayan gözlemlerin de sapan olarak tespit edilmeleri söz konusu olabilmektedir. Bu durum, *Tatlıdil (1981)* tarafından "sürüklenme etkisi" olarak tanımlanmaktadır.

Zaman serilerinde sapan değerlerin analizi ile ilgili ilk çalışmalar *Fox* (1972) tarafından yapılmıştır. Bu çalışmalarda sadece AR modeller için geçerli olan ve olabilirlik oran kriteri adı verilen tarama yöntemi geliştirilmiştir. Daha sonra *Chang* (1983) ve *Chang ve Tiao* (1983), *Fox*' un test yöntemine eklemeler yaparak ARIMA modeller için geçerli olan bir yöntem geliştirmiş ve bunu ardışık yöntem (iterative procedure) olarak isimlendirmişlerdir. Simülasyon çalışmaları, söz konusu yöntemin diğer sapan değer tarama yöntemleri olan ve kaynaklarda LS (Least Square), M-H (Huber-type "down-weighting"), M-B (Bisquare-type "down-weighting"), GM-H (Generalized Huber-type), GM-B (Generalized Bisquare-type) olarak bilinen yöntemlere oranla sapan değer tespit gücü bakımından çok daha üstün ve tutarlı olduğunu göstermiştir. Bu yöntemlere ilişkin simülasyon tabloları bu bölümün sonunda "Yöntemlerin Performans Analizi" başlığı ile verilmektedir (*Chang, Tiao ve Chen, 1988*).

Zaman serilerinde ortaya çıkması muhtemel iki tip sapan değer bulunmaktadır. Birinci tip sapan değer (Additive Outlier (AO)), tek bir gözlemden veya kayıta yapılan büyük bir hata sonucu ortaya çıkar ve hata tek bir gözlemi etkiler. İkinci tip sapan değer (Innovational Outlier (IO)) durumunda birden çok gözlem etkilenir (*Fox, 1972*).

Bir zaman serisi içinde birinci tip sapan değer'in (AO) bulunması, ARMA katsayıları ve σ^2 'nin tahmininde ikinci tip sapan değerlere oranla daha yanlış tahminler elde edilmesine yol açtığı *Ljung* (1993) tarafından belirtilmektedir.

Ayrıca birinci tip sapan değer (AO), bir büyük hata kaynağı olarak düşünülmelidir. Çünkü sadece sapan değer'in bulunduğu T.'inci gözlem etkilenmekte ve parametre tahminlerine olan etki daha büyük olabilmektedir. Halbuki ikinci tip sapan değer durumunda T.'inci pozisyonda meydana gelen etki, kendinden sonraki gözlemlerde de azalarak görülmektedir.

Diğer yandan *Booth* (1987), seriyi oluşturan veriler arasında otokorelasyonun bulunması durumunda, birinci tip sapan değerlerin tespit edilebilme olasılığının önemli ölçüde arttığını saptamıştır.

Zaman serilerinde sapan değer analizi sürekli üretimin yapıldığı kalite kontrol süreçlerinde ortaya çıkan problemlerin çözümünde de kullanılmaktadır. Yapılan üretimle ilgili olarak elde edilen gözlemlerde birinci tip sapan değer elde edilirse, bu durum, üretim faktörlerinin birinde ortaya çıkan aksaklıkla açıklanabilir. Kalite kontrolde "bir kereye özgü hatalı üretim" şeklinde ifade edilen böyle bir durum için muhtemelen hammaddenin sağlıklı ortamlarda muhafaza edilmemesi, üretimin yapılmasını sağlayan makinalarda önceden tahmin edilemeyen bir arızanın meydana gelmesi, işçilerden birinin işini aksatması, malzeme taşıma sistemlerinde bir aksaklığın meydana gelmesi gibi hata kaynaklarından birinin süreci etkilemesi söz konusu olabilir (Bayhan, 1992).

İkinci tip sapan değerinin ortaya çıkması durumu, kalite kontrolde "bir kereye özgü kötü yığın" kullanılmış olabileceği şeklinde yorumlanmaktadır. Bir sonraki üretim sürecinde girdilerde oluşan bozuk hammadde probleminin çözülmesi ile, kötü yığın ortadan kaldırılabılır ve dolayısıyla ikinci tip sapan değer sorununa bir çözüm getirilebilir.

3.4.1. Müdahale Analiz Yöntemi

Bir veri setinde beklenmeyen sıradışı gözlemler sıkça karşılaşılan veri analizi problemleri olarak karşımıza çıkmaktadır. Zaman serisi analizi de buna dahildir. Ekonomik anlamda grevler, savaşlar, ithalat ve ihracat dengesizlikleri gibi önceden tahmin edilmeyen dalgalanmalar sonucunda bazı gözlemler sapan olabilir (Chang, Tiao ve Chen, 1988).

x_t , (p, d, q) parametrelili ARIMA modele ilişkin bir istatistiksel süreç olsun.

$$\phi(B)\alpha(B)x_t = \theta(B)a_t \quad (3.1)$$

Burada p, AR modelin q, MA modelin parametresini d fark sayısını ifade etmektedir. Ayrıca B, $Bx_t = x_{t-1}$ biçiminde tanımlanan geriye öteleme işlecini, $\phi(B) = (1 - \phi_1 B - \dots - \phi_p B^p)$, ve $\theta(B) = (1 - \theta_1 B - \dots - \theta_q B^q)$ polinomları, kökleri birim çemberin dışında kalan iki polinomu; $\alpha(B) = (1 - B)^{d_1} (1 - B^s)^{d_2}$; $d = d_1 + s d_2$;

$\{a_t\} \approx N(0, \sigma_a^2)$ eşitlikleri tanımlamaktadır. Burada d_1 , $\alpha(B)$ fonksiyonunda $(1-B)$ ifadesinin kuvveti ve fark parametresi, sd_2 , ikinci çarpan olarak $(1-B)$ ifadesinin kuvvetini ve mevsimsel fark parametresi, d ise, bu iki fark parametresinin toplamını göstermektedir.

Zaman serisindeki dışsal bir müdahalenin etkisi,

$$z_t = \frac{\omega(B)}{\beta(B)} \zeta_t^r + x_t \quad (3.2)$$

biçiminde ifade edilir. x_t , (3.1) modelinde tanımlı stokastik süreci ifade etmektedir.

$$\begin{aligned} \zeta_t^r &= 1 \quad t = T \\ \zeta_t^r &= 0 \quad \text{Aksi durumda} \end{aligned}$$

müdahalenin oluş zamanını ifade etmektedir. $\omega(B) = (\omega_0 - \omega_1 B - \dots - \omega_s B^s)$ ve $\beta(B) = (\beta_0 - \beta_1 B - \dots - \beta_r B^r)$ iki polinomu ifade etmektedir. $\frac{\omega(B)}{\beta(B)}$ oranı ise müdahalenin dinamik davranışını ifade etmektedir. Burada r , $\beta(B)$ polinomuna ilişkin indis ve kuvveti, s ise, $\omega(B)$ polinomuna ilişkin indis ve kuvveti ifade etmektedir.

Gutman ve Tiao (1978), Miller (1980) ve Chang (1982) yaptığı tespitlere göre müdahale etkisi; ARMA, otokorelasyon ve kısmi otokorelasyon parametrelerini ciddi bir şekilde etkilemektedir. Müdahaleleri tanımlamak, etkilerini gidermek için seriye uygun model denklemini ifade etmek son derece önemlidir (*Chang Tiao ve Chen, 1988*).

Fox (1972), Denby ve Martin (1979), Abraham ve Box (1979), dikkatleri aşağıda tanımlanacak iki müdahale modeli üzerinde toplamışlardır. Bu müdahale modelleri İkinci Tip Sapan Değer (Innovational Outlier (IO)) ve Birinci Tip Sapan Değer (Additive Outlier (AO)) biçiminde isimlendirilmişlerdir.

IO için model;

$$z_t = x_t + \frac{\theta(B)}{\phi(B)\alpha(B)} \omega_{\zeta_t}^T \quad (3.3)$$

AO için model;

$$z_t = x_t + \omega_{\zeta_t}^T \quad (3.4)$$

(3.3) ve (3.4) modeli, a_t 'ler cinsinden;

$$(IO) \quad z_t = \frac{\theta(B)}{\phi(B)\alpha(B)} [a_t + \omega_{\zeta_t}^T] \quad (3.5)$$

$$(AO) \quad z_t = \frac{\theta(B)}{\phi(B)\alpha(B)} a_t + \omega_{\zeta_t}^T \quad (3.6)$$

biçiminde ifade edilmektedir. $\theta(B)$, $\phi(B)$, ve $\alpha(B)$ fonksiyonları (3.1) modelinde tanımlı fonksiyonlardır. Birinci ve ikinci tip sapan değer tiplerini test eden ve olabilirlik oran kriteri adı verilen test yöntemi aşağıda tanıtılmaktadır.

i) ARMA Parametreleri ve σ_a^2 Bilindiğinde :

(3.29) ve (3.30)' teki ω etkisini tahminleyen IO ve AO eşitlikleri gözönüne alındığında zaman serisi parametrelerinin tümü x_t sürecinin varyansı olan σ_a^2 'ya eşittir. Burada J , $p+d+q$ 'den büyük bir tamsayı, ve $1 \leq T \leq n$ ve T, sapan değer ortaya çıktığı pozisyonu ifade etmektedir.

$\pi(B) = \phi(B)\alpha(B)/\theta(B) = 1 - \pi_1 B - \pi_2 B^2 - \dots$ olsun. $\theta(B)$ 'nin kökleri birim çemberin dışında olduğundan π_j 'ler j 'den J 'ye kadar pratikte sıfır değerine yaklaşır.

$$\pi_1 = \phi + (1 - \theta), \quad \pi_2 = (\theta - \phi)(1 - \theta), \quad \pi_j = (\theta - \phi)(1 - \theta)\theta^{j-2} \quad j \geq 3$$

$$e_t = \pi(B)z_t \quad t = 1, \dots, n \text{ olsun.}$$

Bu durumda (3.5) ve (3.6) eşitlikleri sırasıyla;

$$(IO) \quad e_t = \omega \zeta_t^T + a_t \quad (3.7)$$

$$(AO) \quad e_t = \omega \pi(B)\zeta_t^T + a_t \quad (3.8)$$

biçiminde yazılabilir.

Diğer bir deyişle (3.7)'de T noktasındaki hata ile ifade edilen etki, (3.8)'de ise T ve sonraki pozisyonlarda bulunan hata terimleri e_T, e_{T+1}, \dots üzerine dağılan etki anlatılmak istenmektedir.

$$X_t = \pi(B)\zeta_t^T \quad (3.9)$$

$$X_t = 0 \quad t < T$$

$$X_t = 1 \quad t = T$$

En küçük kareler teorisinden, modeldeki ω etkisinin tahmin edicileri aşağıda (3.10) ve (3.11)'de ifade edilmiştir.

$$(IO) \quad \hat{\omega}_I = e_T \quad (3.10)$$

$$(AO) \quad \begin{aligned} \hat{\omega}_A &= \rho^2 \pi(F)e_T \\ &= \rho^2 (1 - \pi_1 F - \pi_2 F^2 - \dots - \pi_{n-T} F^{n-T})e_T \end{aligned} \quad (3.11)$$

Burada $\rho^2 = (1 + \pi_1^2 + \pi_2^2 + \dots + \pi_{n-T}^2)^{-1}$ ve $F, Fe_t = e_{t+1}$ eşitliği ile ifade edilen ileriye öteleme işlecidir. Bunun sonucu olarak IO' nun T noktasındaki etkisi bu noktadaki hata terimi olan e_T ile, AO'nun etkisi ise e_T, e_{T+1}, \dots terimlerinin doğrusal kombinasyonundan tahmin edilebilir. z_t , gözlem terimleri cinsinden (3.11)'deki ω_A tahmini,

$$\hat{\omega}_A = \rho^2 \pi(F)\pi(B)z_t \quad (3.12)$$

biçiminde yazılabilir. Sözkonusu tahmin edicilerin varyansı;

$$\text{var}(\hat{\omega}_I) = \sigma_a^2, \quad \text{var}(\hat{\omega}_A) = \rho^2 \sigma_a^2 \quad (3.13)$$

biçiminde elde edilir. $\rho^2 \leq 1$ olduğundan ω_A 'nın varyansı, $\hat{\omega}_I$ 'dan küçük veya eşittir. Bazı durumlarda ise σ_a^2 'dan daha küçük olabilir. Örneğin x_t süreci birinci dereceden MA modeli ile temsil ediliyorsa, $\hat{\omega}_A$ için varyans, eğer sapan değer serinin sonuna yakın bir pozisyonda değil ise yaklaşık olarak, $\sigma_a^2(1 - \theta^2)$ değeri kadardır (*Chang, Tiao ve Chen, 1988*).

Hipotez Testleri

(3.5) ve (3.6) no'lu eşitlikler ile tanımlanan sapan değer bulunup bulunmadıklarına ilişkin testler aşağıdaki gibi kurulabilir. Burada sıfır hipotezi çalışılan seride sapan değer bulunmadığı, alternatif hipotez, sapan değer bulunduğu varsayımı üzerine kurulmaktadır.

IO sapan değer için

Hipotezler ;

$$\begin{aligned} H_0: \omega &= 0 \\ H_1: \omega &\neq 0 \end{aligned} \quad \text{ve Test İstatistiği } \lambda_{1,T} = \frac{\hat{\omega}_I}{\sigma_a} \quad (3.14)$$

AO sapan değeri için

Hipotezler

$$\begin{aligned} H_0: \omega &= 0 \\ H_1: \omega &\neq 0 \end{aligned} \quad \text{ve Test İstatistiği} \quad \lambda_{2,T} = \frac{\hat{\omega}_A}{\rho \sigma_a} \quad (3.15)$$

IO veya AO Sapan değeri için

Hipotezler ;

$$\begin{aligned} H_0: IO \\ H_1: AO \end{aligned} \quad \text{ve Test İstatistiği} \quad \lambda_{3,T} = [\rho^{-2} \hat{\omega}_A^2 - \hat{\omega}_I^2] / [2\sigma_a^2(1-\rho^2)^{1/2}] \quad (3.16)$$

gibidir.

Bu hipotezde H_0 hipotezinin rededilememesi durumunda ikinci tip sapan değeri, aksi durumda birinci tip sapan değeri bulunduğu sonucuna ulaşılmış olur. H_0 hipotezi altında kullanılan $\lambda_{1,T}$ ve $\lambda_{2,T}$ test istatistikleri standart normal dağılışa sahiptir.

$\lambda_{3,T}$ 'nin dağılışını elde edebilmek için u ve v parametreleri ;

$$u = [\rho^{-1} \hat{\omega}_A + \hat{\omega}_I] / [\sigma_a [1 + \rho]^{1/2}] \quad (3.17)$$

ve

$$v = [\rho^{-1} \hat{\omega}_A - \hat{\omega}_I] / [\sigma_a [1 - \rho]^{1/2}].$$

olsun.

$\lambda_{3,T} = uv$ 'dir. (u, v) ' vektörü, (m_1, m_2) ' ortalama ve $I\sigma^2$ varyansı ile iki değişkenli normal dağılışıdır. Burada $m_1 = \omega [(1 + \rho) / (2\sigma_a^2)]^{1/2}$, $m_2 = -\omega (1 - \rho) / (2\rho\sigma_a^2)^{1/2}$, H_0 hipotezi altında $m_1 = \omega [(1 + \rho) / (2\sigma_a^2)]^{1/2}$, H_1 hipotezi altında da $m_2 = -\omega (1 - \rho) / (2\rho\sigma_a^2)^{1/2}$ eşitlikleri ifade edilmiştir. $\lambda_{3,T}$ dağılışını iki parametre ile ifade edilmektedir. Bunlar; ρ ve ω / σ_a oranıdır. *Craig (1936)*' e göre iki bağımsız

normal dağılışa sahip değişkenin çarpımlarının dağılışı, Bessel fonksiyonlarının bir ifadesi olarak açıklanabilir. H_0 hipotezi altında $\lambda_{3,T}$ 'nin dağılışının yüzde değerleri ω / σ_a ve ρ değerleri için tablolatırılmıştır (Chang, Tiao ve Chen, 1988).

Olabilirlik oran kriteri için test ölçütleri aşağıda ifade edildiği gibidir:

$$(IO) \max_{t=1, \dots, n} |\lambda_{1,T}| \quad (AO) \max_{t=1, \dots, n} |\lambda_{2,T}| \quad (3.18)$$

ii) ARMA Parametreleri ve σ_a^2 Bilinmiyorsa :

Uygulamada ARMA parametreleri ve σ_a^2 bilinmeyebilir. Bu durumda IO veya AO etkileri ile birlikte parametre tahmini olabilirlik fonksiyonu $(\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \omega, \sigma_a^2)$ 'nin maximizasyonu ile elde edilebilir. Bu tahminlere bağlı olarak olabilirlik oranları, (3.16) eşitliğinde ifade edildiği gibi hipotez testleri ile elde edilir ve pozisyonu bilinmeyen sapan değerleri tespit eden yöntemle de sapan değer tarama işlemi yapılır. Parametre tahminleri için kullanılan en çok olabilirlik tahminleyiciler, olabilirlik oranları ile eşdeğer değildir. Bunun aksine genelde karmaşık ifadeleri minimize eder. Çünkü, genel ARMA modelleri ve AO modeli tabiatı gereği doğrusal değildir. Bu bağlamda olabilirlik oran testlerinin uygulanabilir olmadığı ve sapan değerlerin ortaya çıkma anı bilinmediği durumlarda algoritmalara gereksinim

duyulduğu bilinmektedir. $\hat{\phi}_1, \dots, \hat{\phi}_p$; ve $\hat{\theta}_1, \dots, \hat{\theta}_q$; ve $\hat{\sigma}_a^2$ tahmini parametre değerleri,

$\hat{\phi}_1, \dots, \hat{\phi}_p$; ve $\hat{\theta}_1, \dots, \hat{\theta}_q$; ve $\hat{\sigma}_a^2$ parametrelerinin sapan değer etkisi içermeyen en çok olabilirlik tahminleri olsunlar. Bununla birlikte, \hat{e}_t tahmini model ve

$\hat{\pi}(B) = \hat{\phi}(B)\alpha(B) / \hat{\theta}(B)$ 'den hesaplanmış hata terimini gösterebilir. Şimdi aşağıdaki istatistikleri ele alalım.

$$\hat{\lambda}_{1,T} = \hat{\omega}_I / \hat{\sigma}_a \quad \hat{\lambda}_{2,T} = \hat{\omega}_A / (\hat{\rho} \hat{\sigma}_a) \quad (3.19)$$

$$\hat{\lambda}_{3,T} = \frac{\hat{\rho}^{-2} \hat{\omega}_A - \hat{\omega}_I}{2 \hat{\sigma}_a (1 - \hat{\rho})^{1/2}} \quad (3.20)$$

$$\hat{\omega}_I = \hat{e}_T, \quad (3.21)$$

$$\hat{\omega}_A = \hat{\rho} (1 - \pi_1 F - \pi_2 F^2 - \dots - \pi_{n-T} F^{n-T}) \hat{e}_T \quad (3.22)$$

$$\hat{\rho} = (1 + \pi_1 + \pi_2 + \dots + \pi_{n-T})^{-1} \quad (3.23)$$

$\hat{\lambda}_{1,T}, \hat{\lambda}_{2,T}$ ve $\hat{\lambda}_{3,T}$ tahminlerinin beklenen değerleri, asimptotik olarak $\lambda_{1,T}, \lambda_{2,T}$ ve $\lambda_{3,T}$ tahmin edicilerine eşittir. Pozisyonu bilinmeyen IO ve AO sapan değerlerini tespit için aşağıdaki yöntemlerle test edilirler.

$$\eta_{IO} = \max_{t=1, \dots, n} |\hat{\lambda}_{1,T}| > C \quad (\text{IO}) \quad (3.24)$$

$$\eta_{AO} = \max_{t=1, \dots, n} |\hat{\lambda}_{2,T}| > C \quad (\text{AO}) \quad (3.25)$$

Burada C, yüksek duyarlılık için 3.0, orta duyarlılık için 3.5 ve düşük duyarlılık için 4.0 değerleri alabilen, Monte-Carlo simülasyonları ile edilmiş sapan değer ölçütü olarak kullanılan bir sabittir. Seride sapan değerleri (IO ve AO) birbirinden ayırdetmek için kullanılması gerekli basit kural;

$$\left| \hat{\lambda}_{1,T} \right| > \left| \hat{\lambda}_{2,T} \right| \quad \text{ise ikinci tip sapan değer,}$$

$$\left| \hat{\lambda}_{1,T} \right| \leq \left| \hat{\lambda}_{2,T} \right| \quad \text{ise}$$

birinci tip sapan değer şeklindedir (Chang, Tiao ve Chen, 1988).

- **Birinci Tip Sapan Değer (Additive Outlier (AO))**

Tek bir gözlemi etkileyen, bir gözlem veya kayıttaki yapılan hata sonucu ortaya çıkan sapan değer tipidir. Additive Outlier(AO) olarak bilinen sapan değer aşağıda tanımlanan model ile tanımlanmaktadır (Fox, 1972).

$$y_t = z_t + \delta x_t \quad (3.26)$$

Burada y_t gözlem değeri, δ , sapan değerinin büyüklüğü, x_t , T anında 1, diğerleri için 0 değerini alan yapay değişkeni ifade etmektedir.

- **İkinci Tip Sapan Değer (Innovation Outlier(IO))**

Bulunduğu pozisyondan itibaren, kendinden sonraki gözlemleri de etkileyen sapan değer tipidir. Gözlem sapan değeri olarak bilinir (Fox, 1972).

Literatürde Innovation Outlier (IO) olarak tanımlanan model, aşağıdaki gibi gösterilmektedir;

$$y_t = \frac{\theta(B)}{\phi(B)}(e_t + \delta x_t) \quad (3.27)$$

$$\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p,$$
$$\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q.$$

e_t , $N(0, \sigma^2)$ dağılıma sahip hata terimini ifade etmektedir. (3.27)'de yer alan diğer terimler (3.1)'de tanımlandığı gibidir. Bununla birlikte, birinci tip sapan değer, T anındaki büyük bir hatayı ifade ederken, ikinci tip sapan değer, T anında beklenmeyen şok bir etkiyi ifade etmektedir (Ljung, 1993).

Zaman serilerinde sapan değer analizi yapmak amacıyla kullanılan yöntemlere ilişkin Chang Tiao ve Chen (1988) tarafından yapılan simülasyon çalışmaları aşağıdaki gibidir. Sözkonusu simülasyon çalışması AR(1) modeli için yapılmış olup, örneklem büyüklüğü ($n= 50$) olarak kabul edilmiştir. Simülasyon çalışmaları seri içerisinde tek, çift ve üç sapan değerinin bulunması durumuna göre tasarlanmıştır. Seride tek sapan

değer durumu için 26'nci pozisyona ikinci tip sapan değer, yine 26'nci pozisyona birinci tip sapan değer, seride iki sapan değer bulunduğu durumlarda 17 ve 34'üncü pozisyonlara birinci ve ikinci tip sapan değer, seride üç sapan değer bulunduğu durumlar için 14, 26 ve 38'inci pozisyonlara birinci ve ikinci tip sapan değer, ayrıca pozisyonların karışık olduğu durumlar için seri içerisinde sapan değerler yerleştirilmek suretiyle, tarama yöntemlerine ilişkin etkinlik testleri yapılmıştır. Bu testler için ölçüt değer olarak; ortalama, standart sapma ve hata kareler ortalamaları elde edilmiştir (*Chang, Tiao ve Chen, 1988*).

Simülasyon yöntemleri ile etkinlikleri test edilen ve kaynaklarda orijinal isimleri ile anılan yöntemler; LS (Least Squares), M-H (Huber-type, "down-weighting"), M-B (Bisquare-type, "down weighting"), GM-H (Generalized Huber-type), GM-B (Generalized, Bisquare-type) ve Ardışık Yöntem olarak bilinmektedir. Söz konusu yöntemlerin etkinlikleri hata kareler ortalaması bazında ifade edilmektedir.

3.4.2. Ardışık Yöntem

Ardışık Yöntem bir sapan değer tespit yöntemidir. Ardışık Yönteme göre süreçten elde edilen gözlemler setinde sapan değer olup olmadığını araştırmak için önce gözlemlerin bir zaman serisi oluşturduğu düşünülerek serinin uyum gösterdiği istatistiksel modelin parametre tahminleri elde edilir. Bu adımdan sonra seride sapan değer aranır. Seride sapan değer varsa bunların etkileri giderilir. Hiç sapan değer kalmayınca kadar bu işlem tekrarlanır. Seride sapan değere rastlanmadığı takdirde o ana kadar hangi gözlemlere karşı gelen noktalarda sapan değere rastlanmışsa o gözlem değerleri için yeni değerler hesaplanır. Sapan değerlerin bulunduğu gözlem değerleri için yeni değerler hesaplanıp eskileri ile değiştirildikten sonra yeni seri için parametre kestirimleri elde edilir ve sapan değer tarama işlemi tekrar edilir. Bu aşamada da sapan değere rastlanmazsa işleme son verilir. Ardışık Yöntemin algoritması aşağıda aşamalar biçiminde ifade edilmiştir :

1. Aşama : Tahmini modelden tahmini hata terimleri elde edilir. (\hat{e}_t) ve σ_a^2 'nin tahmini olarak $\hat{\sigma}_a^2 = n^{-1} \sum_{t=1}^n \hat{e}_t^2$ hesaplanır. σ_a^2 'nin alternatif tahminleri, hataların mutlak değerlerinin medyanına bağlı olarak elde edilebilir.

2. Aşama : $\hat{\lambda}_{1,T}$ ve $\hat{\lambda}_{2,T}$ hesaplanarak, $\eta_t = \max\{\hat{\lambda}_{1,T}, \hat{\lambda}_{2,T}\}$ $t = 1, \dots, n$ elde edilir.

Eğer, $|\hat{\lambda}_{1,T}| > C$ (C burada kullanılan ölçüt değeri) ise, T noktasında bir ikinci tip sapan değer (IO) bulunması olasıdır. İkinci tip sapan değeri etkisi $(\hat{\omega}_I)$, (3.10)'da ifade edildiği gibi elde edilir. Bu etki giderilerek T noktası için yeni bir hata terimi elde edilir.

$(\hat{e}_T = \hat{e}_T - \hat{\omega}_I = 0)$ Eğer $\eta_t = \max|\hat{\lambda}_{2,T}| > C$ ise T noktasında birinci tip sapan değeri (AO)'nun bulunması olasılığından söz edilebilir. Birinci tip sapan değeri etkisi (3.11)'de ifade edildiği gibi, T noktasına yeni bir hata teriminin hesaplanmasıyla giderilebilir.

$(\hat{e}_T = \hat{e}_T - \hat{\omega}_A \pi(B) \zeta_t^T \quad t \geq T)$ Daha sonra değiştirilmiş hata terimlerinden yeni $\hat{\sigma}_a^2$ elde edilir.

3. Aşama : Eğer ikinci aşamada bir IO veya AO elde edilmiş ise, zaman serisi parametrelerinin son tahminlerine, değiştirilmiş hata terimlerine ve $\hat{\sigma}_a^2$ 'ya bağlı kalınarak yeni $\hat{\lambda}_{1,T}$ ve $\hat{\lambda}_{2,T}$ hesaplanır ve ikinci aşama tekrar edilir.

4. Aşama : 2 ve 3 nolu aşamalar sapan değerler sona erinceye kadar tekrar edilir (Chang, Tiao ve Chen, 1988).

3.4.3. Sapan Değer Tespit Ölçütleri ve Olasılığı

Sapan değer analizi çalışmalarında, sapan değerlerin tespit edilmesine yönelik çeşitli faktörlerden söz edilir. Ancak ne sebeple ortaya çıkarsa çıksın sapan değerleri tespit işlemi ile birlikte bir çok cevaplanması gerekli soru vardır. Doğrudan hata terimleri ile ilişkili olan sapan değer analizi çalışmalarında eldeki gözlem değerlerine uygun modeli belirlemek son derece önemlidir. Verilerin genel eğilimini belirleme işlemi için seri uzunluğunun mümkün olduğu kadar büyük ölçekli olması genel eğilimdir. Bu büyüklüğün hacmi kesin olarak tespit edilememiştir. Sapan olan gözlemin ne kadar sapsmış olduğu ise cevaplanması gerekli bir başka sorun gibi görünmektedir. Sapan gözlemin sapma derecesi, gözlemin istatistiksel ölçütlere göre sapan olup olmadığı ile belirlenmektedir. Sapan olan gözlemin rasgele mi yoksa kişilerden veya ölçü aletlerinden mi kaynaklandığı bir başka problemidir. Yukarıda sözü edilen unsurları bir araya getirip çözmek amacıyla sapan değerlerin tespitinde model, sapan değer büyüklüğü, seri genişliği, sapan değer tipleri, ölçüt değer duyarlılığı ve seri içerisinde bulunan sapan değer adedinin etkileri tespit edilmeye çalışılacaktır. Ayrıca sapan değerlerin tespiti aşamasında kullanılan tarama yöntemlerinin güçlerinin istatistiksel açıdan farklı olup olmadığı belirlenecektir. Yapılacak varyans analizi çalışmalarında kullanılan veriler *Chang, Tiao ve Chen (1988)* tarafından simülasyon çalışmaları ile elde edilmiştir. Söz konusu veriler sapan değer tespit olasılıkları olup şu şekilde elde edilmiştir:

Zaman serisi modeli (AR, MA), sapan değer tipi (AO, IO ve AO+IO), seri uzunluğu (50, 100, 150), sapan değer büyüklüğü (3σ , 5σ), ölçüt değer duyarlılığı (3.00, 3.50, 4.00) gibi özellikleri önceden belirlenen seriler için simülasyon yöntemi ile veriler türetilmektedir. Türetilen verilerle oluşturulan serilerin belirlenen pozisyonlarına sapan değerler yerleştirilmekte ve bu işlem 1000 kez tekrar edilmektedir. Tekrar işlemlerinin her biri için seriyi oluşturan veriler farklı türetildiklerinden, tarama yöntemleri aynı özelliklere sahip 1000 farklı seri üzerinde sapan değer tarama işlemi gerçekleştirilmektedir. Sapan değer tespit olasılığı, tespit sayısının 1000'e bölünmesi ile elde edilen değerdir.

Tek sapan değer durumu için AR(1), $\phi = .6$ ve $\sigma_a^2 = 1$, MA(1), $\theta = .6$ ve $\sigma_a^2 = 1$ değerleri, seri genişlikleri; n=50, n=100 ve n=150 için, $\omega = 3\sigma_a$ ve $\omega = 5\sigma_a$ sapan değer etkileri gözönünde bulundurularak yapılan simülasyon çalışmaları ele alınmıştır. Bu çalışmalarda sapan değerler, gözlemsel periyotların ortasına set edilmiştir. Yani, n=50 için T=26, n=100 için T=51 ve n=150 için T=76'ya set edilmişlerdir.

Cift sapan değer durumu için AR(1) $\phi = .6$ ve $\sigma_a^2 = 1$, parametreleri ve sapan değerlerin bulunması muhtemel üç farklı durum (sapan değerlerden biri AO, diğeri IO, sapan değerlerin ikisi de AO veya sapan değerlerin ikisi de IO) gözönünde bulundurularak n=50, 100 ve 150 seri genişlikleri, $\omega = 3\sigma_a$ ve $\omega = 5\sigma_a$ sapan değer etkileri için yapılan simülasyon çalışmaları kullanılmıştır. Sapan değerler, 50 seri genişliği üçe bölünerek ilk ve ikinci periyotlara yerleştirilmişlerdir. Yani sapan değer pozisyonları $T_1 = 17$ ve $T_2 = 34$ olmuştur. Ölçümü 100 olan bir seri için sapan değer pozisyonları $T_1 = 34$ ve $T_2 = 66$, ölçümü 150 olan bir seri için ise sapan değer pozisyonları, $T_1 = 51$ ve $T_2 = 101$ olarak alınmıştır.

Yukarıdaki ön tasarımlar özel ölçütlere bağlı elde edilmiş zaman serisi gözlemleri olarak düşünülmelidir. Ayrıca, modelin bilindiği ve örnekleme de sanki sapan değerlerin bulunmadığı varsayımı yapılmıştır. Bunun yanında hataların yaratılması ve zaman serisi parametrelerinin tahmini, *Hillmer ve Tiao (1979)* tarafından geliştirilen tam olabilirlik tahmin yöntemiyle elde edilmiştir. Yukarıda belirtilen sapan değer durumlarına ve hatalara bağlı olarak C=4.0 (düşük duyarlılık) kritik değeri için Tablo-3.1, C=3.5 (orta duyarlılık) kritik değeri için Tablo-3.2 ve C=3.0 (yüksek duyarlılık) kritik değeri için ise Tablo-3.3'te verilen sonuçlar elde edilmiştir. Alternatif ölçütlerin önemliliğini belirtmek için bu sonuçlar varyans analizine tabi tutulmuşlardır.

Tablo-3.1, C=4.0, 1000 Tekrarlı, Sapan Değer Tespit Olasılıkları.

	$\omega = 3\sigma_a$ (Sapan Değer Etkisi)			$\omega = 5\sigma_a$ (Sapan Değer Etkisi)		
	n=50	n=100	n=150	n=50	n=100	n=150
AR 1 AO	0.72	0.80	0.81	0.91	0.90	0.92
AR 1 IO	0.88	0.87	0.87	0.95	0.93	0.91
MA 1 AO	0.70	0.82	0.84	0.83	0.90	0.93
MA 1 IO	0.97	0.92	0.88	0.99	0.96	0.97
AR 2 AO	0.38	0.54	0.52	0.53	0.76	0.80
Birinci Sapan Değer	0.65	0.74	0.81	0.76	0.86	0.92
İkinci Sapan Değer	0.64	0.73	0.78	0.74	0.88	0.88
AR 2 IO	0.58	0.67	0.55	0.82	0.86	0.84
Birinci Sapan Değer	0.84	0.80	0.85	0.91	0.94	0.91
İkinci Sapan Değer	0.77	0.88	0.83	0.92	0.91	0.92
AR, IO AO	0.65	0.63	0.69	0.86	0.87	0.88
Birinci Sapan Değer	0.90	0.86	0.84	0.98	0.95	0.93
İkinci Sapan Değer	0.72	0.81	0.81	0.88	0.92	0.93

Tablo-3.2, C=3.5, 1000 Tekrarlı, Sapan Değer Tespit Olasılıkları.

	$\omega = 3\sigma_a$ (Sapan Değer Etkisi)			$\omega = 5\sigma_a$ (Sapan Değer Etkisi)		
	n=50	n=100	n=150	n=50	n=100	n=150
AR 1 AO	0.75	0.80	0.82	0.86	0.90	0.92
AR 1 IO	0.85	0.88	0.81	0.94	0.93	0.90
MA 1 AO	0.75	0.83	0.84	0.84	0.91	0.93
MA 1 IO	0.94	0.90	0.89	0.98	0.96	0.96
AR 2 AO	0.44	0.54	0.62	0.58	0.76	0.80
Birinci Sapan Değer	0.70	0.75	0.81	0.78	0.86	0.92
İkinci Sapan Değer	0.70	0.76	0.80	0.77	0.88	0.88
AR 2 IO	0.64	0.69	0.65	0.86	0.87	0.88
Birinci Sapan Değer	0.88	0.84	0.84	0.97	0.95	0.93
İkinci Sapan Değer	0.75	0.82	0.79	0.88	0.92	0.93
AR, IO AO	0.65	0.63	0.69	0.86	0.87	0.88
Birinci Sapan Değer	0.90	0.86	0.84	0.98	0.95	0.93
İkinci Sapan Değer	0.72	0.81	0.81	0.88	0.92	0.93

Tablo-3.3, C=3.0, 1000 Tekrarlı, Sapan Değer Tespit Olasılıkları.

	$\omega = 3\sigma_a$ (Sapan Değer Etkisi)			$\omega = 5\sigma_a$ (Sapan Değer Etkisi)		
	n=50	n=100	n=150	n=50	n=100	n=150
AR 1 AO	0.75	0.81	0.81	0.86	0.90	0.92
AR 1 IO	0.83	0.82	0.77	0.93	0.93	0.90
MA 1 AO	0.78	0.83	0.85	0.84	0.91	0.93
MA 1 IO	0.91	0.87	0.88	0.98	0.96	0.97
AR 2 AO						
	0.50	0.59	0.63	0.60	0.76	0.80
Birinci Sapan Değer	0.74	0.77	0.81	0.79	0.86	0.91
İkinci Sapan Değer	0.72	0.77	0.80	0.77	0.88	0.88
AR 2 IO						
	0.61	0.63	0.65	0.81	0.85	0.82
Birinci Sapan Değer	0.82	0.79	0.85	0.91	0.93	0.90
İkinci Sapan Değer	0.78	0.82	0.80	0.90	0.90	0.91
AR, IO AO						
	0.65	0.68	0.65	0.86	0.86	0.87
Birinci Sapan Değer	0.87	0.82	0.84	0.97	0.94	0.93
İkinci Sapan Değer	0.76	0.83	0.79	0.89	0.91	0.93

Tablo-3.1, Tablo-3.2 ve Tablo-3.3 incelendiğinde, tek sapan değer durumunda tarama yönteminin sapan değer tespit olasılığı çok yüksek çıkmaktadır. Seri genişliklerinin büyütülmesi yöntemin gücünü artıran faktör gibi görünmektedir. Bunun yanında sapan değer'in büyüklüğü, teşhis edilmesinde önemli bir etken gibi görünmektedir. Dikkati çeken önemli bir husus; seri içinde sapan değer sayısının bir'den iki'ye çıkması, yöntemin sapan değer tespit gücünde büyük oranda güç kaybına neden olduğudur. Diğer bir nokta; model farklılıklarının (AR, MA) sapan değerleri tespitinde belirleyici olduklarıdır. Tablo-3.1, Tablo-3.2 ve Tablo-3.3 dikkatle incelenecek olursa MA modellerde sapan değerlerin tespit edilmesi daha kolay olmaktadır. Ayrıca, IO olarak tanımlanan "ikinci tip sapan değer", AO olarak tanımlanan "birinci tip sapan değer" e göre daha kolay tespit edilmektedir. Çıkarılması gerekli bir diğer sonuç, MA modellerde IO sapan değerinin tespit olasılığının çok yüksek olduğu şeklindedir. Görsel olarak elde edilen sonuçları istatistiksel testlerle doğrulamak amacıyla, tek ve çift sapan değerler için varyans analizi ile doğrulama gereği vardır. Bu işlemler için, Tablo-3.1, Tablo-3.2 ve Tablo-3.3' ün tekrar ele alınmış ve sözü edilen analiz sonuçlarına Bölüm 4'te yer verilmiştir.

3.4.4. Tarama Yöntemleri için Performans Analizi

Aşağıda verilen tablolarda n=50 ölçümü için tek, çift ve üç sapan değerinde bulunduğu serilerde sapan değer tarama yöntemlerinin performansları; ortalama, standart sapma ve hata kareler ortalamaları bazında ifade edilmektedir. Performans ölçütleri olarak kabul edilen ortalama, standart sapma ve hata kareler ortalaması bilgileri, sapan olarak elde edilen gözlem değerleri yerine yeni gözlem değerlerinin hesaplanıp tekrar analizi ile bulunmaktadır. Burada etkinlik, sapan değer tiplerini doğru tanımlama ve sapan olarak elde edilen değerler yerine, yeni değerlerin etkili bir şekilde hesaplanması esasına dayanmaktadır.

Tek Sapan Değer (IO)

Tablo-3.4, Tek Sapan Değer (IO) İçin Yöntemlerin Performansı.

n=50, T=26 (IO)		PARAMETRE DEĞERLERİ		
Sapan Değer Tarama Yöntemleri	Ortalama	Standart Sapma	Hata Kareler Ortalaması	
LS	0.8688	0.0787	0.0071	
M-H	0.8752	0.0708	0.0056	
M-B	0.8771	0.0703	0.0055	
GM-H	0.8786	0.0711	0.0055	
GM-B	0.8783	0.0717	0.0056	
ARDIŞIK	0.8813	0.0673	0.0049*	

Tek Sapan Değer (AO)

Tablo-3.5, Tek Sapan Değer (AO) İçin Yöntemlerin Performansı.

n=50, T=26 (AO)		PARAMETRE DEĞERLERİ		
Sapan Değer Tarama Yöntemleri	Ortalama	Standart Sapma	Hata Kareler Ortalaması	
LS	0.7686	0.1155	0.0306	
M-H	0.8163	0.1073	0.0185	
M-B	0.8264	0.1137	0.0183	
GM-H	0.8440	0.0938	0.0119	
GM-B	0.8555	0.0894	0.0100	
ARDIŞIK	0.8658	0.0819	0.0079*	

Çift Sapan Değerler (AO)

Tablo-3.6, Çift Sapan Değer (AO) İçin Yöntemlerin Performansı.

n=50, T=17, 34 (AO)		PARAMETRE DEĞERLERİ		
Sapan Değer Tarama Yöntemleri	Ortalama	Standart Sapma	Hata Kareler Ortalaması	
LS	0.7483	0.1189	0.0371	
M-H	0.7927	0.1106	0.0237	
M-B	0.7962	0.1173	0.0245	
GM-H	0.8267	0.0952	0.0144	
GM-B	0.8353	0.0924	0.0127	
ARDIŞIK	0.8601	0.0855	0.0089*	

Çift Sapan Değerler (IO)

Tablo-3.7, Çift Sapan Değerler (IO) İçin Yöntemlerin Performansı.

n=50, T=17, 34 (IO)		PARAMETRE DEĞERLERİ		
Sapan Değer Tarama Yöntemleri	Ortalama	Standart Sapma	Hata Kareler Ortalaması	
LS	0.8678	0.0693	0.0058	
M-H	0.8745	0.0631	0.0046	
M-B	0.8768	0.0629	0.0045	
GM-H	0.8768	0.0645	0.0047	
GM-B	0.8765	0.0698	0.0054	
ARDIŞIK	0.8802	0.0635	0.0044*	

Üç Sapan Değerler (AO)

Tablo-3.8, Üç Sapan Değer (AO) İçin Yöntemlerin Performansı.

n=50, T=14, 26, 38 (AO) PARAMETRE DEĞERLERİ			
Sapan Değer Tarama Yöntemleri	Ortalama	Standart Sapma	Hata Kareler Ortalaması
LS	0.6845	0.1420	0.0665
M-H	0.7360	0.1421	0.0470
M-B	0.7340	0.1507	0.0502
GM-H	0.7932	0.1131	0.0242
GM-B	0.8078	0.1084	0.0202
ARDIŞIK	0.8554	0.1000	0.0120*

Üç Sapan Değerler (IO)

Tablo-3.9, Üç Sapan Değer (IO) İçin Süreçlerinin Performansı.

n=50, T=14, 26, 38 (IO) PARAMETRE DEĞERLERİ			
Sapan Değer Tarama Yöntemleri	Ortalama	Standart Sapma	Hata Kareler Ortalaması
LS	0.7480	0.1016	0.0334
M-H	0.7956	0.0970	0.0203
M-B	0.7961	0.1046	0.0217
GM-H	0.8301	0.0814	0.0115
GM-B	0.8393	0.0787	0.0099
ARDIŞIK	0.8743	0.0675	0.0052*

Üç Sapan Değerler (Karışık)

Tablo-3.10, Üç Sapan Değer İçin Süreçlerin Performansı.

n=50, Pozisyonlar Karışık			
PARAMETRE DEĞERLERİ			
Sapan Değer Tarama Yöntemleri	Ortalama	Standart Sapma	Hata Kareler Ortalaması
LS	0.8110	0.1268	0.0240
M-H	0.8420	0.0914	0.0117
M-B	0.8541	0.0837	0.0091
GM-H	0.8592	0.0756	0.0074
GM-B	0.8621	0.0787	0.0076
ARDIŞIK	0.8688	0.0718	0.0061*

Tarama süreçlerinin performanslarını belirlemek amacıyla *Chang Tiao ve Chen (1988)* tarafından elde edilmiş sonuçlar; yukarıda verildiği üzere 7 ayrı tablo biçiminde sunulmuştur. Sapan değer sayıları esas alınarak hazırlanan tabloların incelenmesi ile ortalama, standart sapma ve hata kareler ortalaması bazında Ardışık Yöntemin üstünlüğü açık bir şekilde görülmektedir. Bu üstünlüğün istatistiksel bakımdan ne derece önemli olduğunu ortaya koymak amacıyla Bölüm 4'de varyans analizi yöntemi ile yeniden test işlemi yapılmıştır.

BÖLÜM 4

4. SAPAN DEĞER TESPİTİNDE FAKTÖR ANA ETKİLERİNİN ÖNEM DERECELERİNİN VARYANS ANALİZİ İLE BELİRLENMESİ VE UYGULAMA

4.1. Tek Sapan Değerlere İlişkin Varyans Analizi Çalışması

Chang Tiao ve Chen (1988) tarafından yapılmış simülasyon sonuçları kullanılarak; sapan değer taramada; Model (M) (AR, MA), sapan değer Büyüklüğü (B) (3σ , 5σ), seri Genişliği (G) (50, 100, 150), ölçüt değer Duyarlılığı (D) (C=3.0, C=3.5, C=4.0) ve sapan değer Türü (T) (AO, IO) gibi faktörlerin ne kadar etkili olduğunu istatistiksel olarak ortaya koymaya yönelik varyans analizi çalışması yapılmıştır. Faktörlerin seçimi, simülasyon sonuçlarına bağlı ve sapan değer çalışmalarında çözüm aranan ölçütler olması ile ilişkili olarak aşağıda belirtilen şekliyle ele alınmıştır :

Faktör Ana Etkileri ;

Sapan Değer Modeli (M),

Sapan Değer Büyüklüğü (B) ,

Seri Genişliği (G),

Ölçüt Değer Duyarlılığı (D),

Sapan Değer Türü (T).

Bu faktörlere ilişkin iki'li etkileşimler;

M*B, M*G, M*D, M*T.

B*G, B*D, B*T.

G*D, G*T.

D*T.

Faktörlere ilişkin üç'lü etkileşimler;

M*B*G, M*B*D, M*B*T, M*G*D, M*G*T, M*D*T.

B*G*D, B*G*T, B*D*T.

G*D*T.

Faktörlere İlişkin dört'lü ve beş'li etkileşimler;

M*B*G*D, M*B*G*T, M*B*D*T, M*G*D*T, B*G*D*T.

M*B*G*D*T gibidir.

Özel seçimli, çok-etkenli (2x2x3x3x2) (2³ x 3²) deney düzeni için model denklemi aşağıdaki gibi tasarlanmıştır :

$$Y_{ijklm} = \mu + M_i + B_j + G_k + D_l + T_m + (MB)_{ij} + (MG)_{ik} + (MD)_{il} + (MT)_{im} + (BG)_{jk} + (BD)_{jl} + (BT)_{jm} + (GD)_{kl} + (GT)_{km} + (DT)_{lm} + (MBG)_{ijk} + (MBD)_{ijl} + (MBT)_{ijm} + (BGD)_{jkl} + (BGT)_{jkm} + (GDT)_{klm} + \varepsilon_{ijklm} \quad (4.1)$$

$$i = 1,2; \quad j = 1,2; \quad k = 1,2,3; \quad l = 1,2,3; \quad m = 1,2.$$

Y_{ijklm} : Simülasyon sonuçları ile elde edilen tek sapan değer tespit olasılıkları.

μ : Genel ortalama.

(4.1) modelinde bazı 3'lü etkileşimler ile 4'lü ve 5'li etkileşimler hatanın tahminlenmesine olanak vermek için sıfır varsayılmıştır. Dolayısıyla söz konusu etkileşimlerin bağımlı değişkeni açıklama oranları hata terimi içerisine dahil edilmiştir.

Tablo-4.1, Tek Sapan Değer Tarama Olasılıkları.

		C=3.0		C=3.5		C=4.0		
		AO	IO	AO	IO	AO	IO	
AR	3σ	50	0.75	0.83	0.75	0.85	0.72	0.88
		100	0.81	0.82	0.80	0.88	0.80	0.87
		150	0.81	0.77	0.82	0.81	0.81	0.87
	5σ	50	0.86	0.93	0.86	0.94	0.91	0.95
		100	0.90	0.93	0.90	0.93	0.90	0.93
		150	0.92	0.90	0.92	0.90	0.92	0.91
MA	3σ	50	0.78	0.91	0.75	0.94	0.70	0.97
		100	0.83	0.87	0.83	0.90	0.82	0.92
		150	0.85	0.88	0.84	0.89	0.84	0.88
	5σ	50	0.84	0.98	0.84	0.98	0.83	0.99
		100	0.91	0.96	0.91	0.96	0.90	0.96
		150	0.93	0.97	0.93	0.96	0.93	0.97

Tablo-4.1'deki olasılık değerleri Tablo-4.2' de tasarım çizelgesi ile birlikte verilmiştir. Y değerleri Chang Tiao ve Chen (1988) tarafından yapılan simülasyon çalışmaları ile elde edilen olasılık değerleridir. Dikkat edileceği gibi tasarım çizelgesi ile verilen eleman sayısı, $i = 1,2$; $j = 1,2$; $k = 1,2,3$; $l = 1,2,3$; $m = 1,2$. seviye sayılarının çarpımları kadardır.

Tablo-4.2, (4.1) Modeline İlişkin Tasarım Çizelgesi.

No	Y	M	B	G	D	T	No	Y	M	B	G	D	T
1	0.75	1	1	1	1	1	37	0.78	2	1	1	1	1
2	0.83	1	1	1	1	2	38	0.91	2	1	1	1	2
3	0.75	1	1	1	2	1	39	0.75	2	1	1	2	1
4	0.85	1	1	1	2	2	40	0.94	2	1	1	2	2
5	0.72	1	1	1	3	1	41	0.70	2	1	1	3	1
6	0.88	1	1	1	3	2	42	0.97	2	1	1	3	2
7	0.81	1	1	2	1	1	43	0.83	2	1	2	1	1
8	0.82	1	1	2	1	2	44	0.87	2	1	2	1	2
9	0.80	1	1	2	2	1	45	0.83	2	1	2	2	1
10	0.88	1	1	2	2	2	46	0.90	2	1	2	2	2
11	0.80	1	1	2	3	1	47	0.82	2	1	2	3	1
12	0.87	1	1	2	3	2	48	0.92	2	1	2	3	2
13	0.81	1	1	3	1	1	49	0.85	2	1	3	1	1
14	0.77	1	1	3	1	2	50	0.88	2	1	3	1	2
15	0.82	1	1	3	2	1	51	0.84	2	1	3	2	1
16	0.81	1	1	3	2	2	52	0.89	2	1	3	2	2
17	0.81	1	1	3	3	1	53	0.84	2	1	3	3	1
18	0.87	1	1	3	3	2	54	0.88	2	1	3	3	2
19	0.86	1	2	1	1	1	55	0.84	2	2	1	1	1
20	0.93	1	2	1	1	2	56	0.98	2	2	1	1	2
21	0.86	1	2	1	2	1	57	0.84	2	2	1	2	1
22	0.94	1	2	1	2	2	58	0.98	2	2	1	2	2
23	0.91	1	2	1	3	1	59	0.83	2	2	1	3	1
24	0.95	1	2	1	3	2	60	0.99	2	2	1	3	2
25	0.90	1	2	2	1	1	61	0.91	2	2	2	1	1
26	0.93	1	2	2	1	2	62	0.96	2	2	2	1	2
27	0.90	1	2	2	2	1	63	0.91	2	2	2	2	1
28	0.93	1	2	2	2	2	64	0.96	2	2	2	2	2
29	0.90	1	2	2	3	1	65	0.90	2	2	2	3	1
30	0.93	1	2	2	3	2	66	0.96	2	2	2	3	2
31	0.92	1	2	3	1	1	67	0.93	2	2	3	1	1
32	0.90	1	2	3	1	2	68	0.97	2	2	3	1	2
33	0.92	1	2	3	2	1	69	0.93	2	2	3	2	1
34	0.90	1	2	3	2	2	70	0.96	2	2	3	2	2
35	0.92	1	2	3	3	1	71	0.93	2	2	3	3	1
36	0.91	1	2	3	3	2	72	0.97	2	2	3	3	2

Tablo-4.2' de verilen veriler, MINITAB istatistik analiz paket programı ile çok etkenli varyans analizine tabi tutulmuş ve Tablo-4.3' te verilen gözlem, tahmin ve hata terimleri elde edilmiştir. Aynı analiz işlemi ile, Tablo-4.4' te verilen varyans analizi sonuçları elde edilmiştir. Tablo-4.2 ile verilen tablonun MINITAB paket programı ile analiz edilme işlemi aşağıda tanımlandığı şekliyle ANOVA komutunun kullanımı ile yapılmaktadır.

- **MINITAB Paket Programı ile Analiz İşlemi :**

```
BATCH
OUTFILE 'SONUC.DAT'
READ C1-C6
0.75 1 1 1 1 1
0.83 1 1 1 1 2
0.75 1 1 1 2 1
0.85 1 1 1 2 2
0.72 1 1 1 3 1
0.88 1 1 1 3 2
0.81 1 1 2 1 1
.....
.....
.....
0.97 2 2 3 3 2
END
```

```
NAME C1 'Y'
NAME C2 'M'
NAME C3 'B'
NAME C4 'G'
NAME C5 'D'
NAME C6 'T'
```

```
ANOVA Y=M B G D T M*B M*G M*D M*T B*G B*D B*T G*D
      G*T D*T M*B*G M*B*D M*B*T B*G*D B*G*T G*D*T
```

Gözlemlere ilişkin tahmin ve hata değerlerini elde etmek için ANOVA komutunu kullanmak ve tabloda ilgili sahalari işaretlemek yeterlidir. Bu işlemler sonucunda özel seçimli, çok etkenli $(2 \times 2 \times 3 \times 3 \times 2)$ $(2^3 \times 3^2)$ deney düzeni için analiz gerçekleştirilmiş olur. Analiz sonuçları Tablo-4.3 ve Tablo-4.4' te verilmiştir.

Tablo-4.3, (4.1) Modeline İlişkin Gözlem Tahmin ve Hata Değerleri.

No	Gözlem	Tahmin	Hata	No	Gözlem	Tahmin	Hata
1	0.75	0.73	0.02	37	0.78	0.77	0.01
2	0.83	0.84	-0.01	38	0.91	0.92	-0.01
3	0.75	0.74	0.01	39	0.75	0.76	-0.01
4	0.85	0.87	-0.02	40	0.94	0.93	0.01
5	0.72	0.72	0.00	41	0.70	0.73	-0.03
6	0.88	0.88	0.00	42	0.97	0.94	0.03
7	0.81	0.80	0.01	43	0.83	0.82	0.01
8	0.82	0.82	0.00	44	0.87	0.89	-0.02
9	0.80	0.81	-0.01	45	0.83	0.82	0.01
10	0.88	0.86	0.02	46	0.90	0.91	-0.01
11	0.80	0.82	-0.02	47	0.82	0.81	-0.01
12	0.87	0.87	0.00	48	0.92	0.91	0.01
13	0.81	0.80	0.01	49	0.85	0.84	0.01
14	0.77	0.79	-0.02	50	0.88	0.87	0.01
15	0.82	0.82	0.00	51	0.84	0.84	0.00
16	0.81	0.87	-0.06	52	0.89	0.88	0.01
17	0.81	0.98	-0.17	53	0.84	0.84	0.00
18	0.87	0.91	-0.04	54	0.88	0.90	-0.02
19	0.86	0.91	-0.05	55	0.84	0.85	-0.01
20	0.93	0.90	0.03	56	0.98	0.96	0.00
21	0.86	0.89	-0.03	57	0.84	0.84	0.00
22	0.94	0.94	0.00	58	0.98	0.97	0.01
23	0.91	0.87	0.04	59	0.83	0.84	-0.01
24	0.95	0.98	-0.03	60	0.99	0.10	-0.01
25	0.90	0.91	-0.01	61	0.91	0.91	0.00
26	0.93	0.91	0.02	62	0.96	0.96	0.00
27	0.90	0.90	0.00	63	0.91	0.90	0.01
28	0.93	0.92	0.01	64	0.96	0.97	-0.01
29	0.90	0.90	0.00	65	0.90	0.89	0.01
30	0.93	0.93	0.00	66	0.96	0.97	-0.01
31	0.92	0.92	0.00	67	0.93	0.94	-0.01
32	0.90	0.89	0.01	68	0.97	0.96	0.01
33	0.92	0.92	0.00	69	0.93	0.93	0.00
34	0.90	0.90	0.00	70	0.96	0.96	0.00
35	0.92	0.92	0.00	71	0.93	0.92	0.01
36	0.91	0.92	-0.01	72	0.97	0.97	0.00

Tablo-4.4, (4.1) Modeline İlişkin Varyans Analizi Sonuçları.

D.K.	S.D.	K.T.	K.O.	F.	p.
M (Model)	1	0.016501	0.016501	41.78	0.000
B (Büyüklik)	1	0.134335	0.134335	340.09	0.000
G (Genişlik)	2	0.006808	0.003404	8.62	0.001
D (Duyarlılık)	2	0.001225	0.000613	1.55	0.229
T (Tür)	1	0.078013	0.078013	197.50	0.000
M*B	1	0.002335	0.002355	5.91	0.021
M*G	2	0.001353	0.000676	1.71	0.198
M*D	2	0.001203	0.000601	1.52	0.235
M*T	1	0.011001	0.011001	27.85	0.000
B*G	2	0.000586	0.000293	0.74	0.485
B*D	2	0.000486	0.000243	0.62	0.547
B*T	1	0.003335	0.003335	8.44	0.007
G*D	4	0.000192	0.000048	0.12	0.974
G*T	2	0.040908	0.020454	51.78	0.000
D*T	2	0.004408	0.002204	5.58	0.009
M*B*G	2	0.0009528	0.0004764	1.21	0.313
M*B*D	2	0.0001194	0.0000597	0.15	0.860
M*B*T	1	0.0000681	0.0000681	0.17	0.681
B*G*D	4	0.0010306	0.0002576	0.65	0.630
B*G*T	2	0.0012194	0.0006097	1.54	0.230
G*D*T	4	0.0004583	0.0001146	0.29	0.882
Hata	30	0.0118500	0.000395		
Toplam	71	0.318388			

Tek sapan değer verileri için yapılan varyans analizi işlemi ve elde edilen varyans analizi tablosu ile sapan değerlerin Model (M), Büyüklik (B), Genişlik (G), Duyarlılık (D) ve Tür (T) etkileri bağımsız olarak, ayrıca iki'li ve bazı üç'lü etkileşimler yönünden incelenmiştir. Buna göre;

- Zaman serisi modeli, (AR, MA) sapan değer tespit edilmesinde önemli bir etken olarak elde edilmiştir. Bu farklılık, yukarıdaki tablodan gözlenebilir.
- Sapan değer büyüklüğü'nün (3σ , 5σ), sapan değer tespitinde istatistiki açıdan önemli olduğu sonucu elde edilmiştir.

- Seri genişliklerinin (50, 100, 150), sapan değer tespitinde önemli bulunduğu ancak, bu farkın 50 ile 100 aralığında önemli, 100 ile 150 aralığında önemsiz olduğu görülmektedir.
- Sapan değer tarama ölçütü C'nin farklı değerlere sahip olması (C=3.0, C=3.5, C=4.0), sapan değerleri tespit yönünden önemsiz bulunmuştur.
- Sapan değer türlerinin (AO, IO), sapan değer belirlemede önemli bir etken olduğu sonucuna varılmıştır.

Tek sapan değer verileri için oluşturulan iki'li etkileşimlere bakıldığında; büyüklük ve tür ile oluşturulan iki'li etkileşimler önemli, diğer iki'li ve modele dahil edilen üç'lü etkileşimler önemsiz bulunmuştur. (4.1) Modeli için elde edilen varyans analiz tablosu incelendiğinde, hata değerinin toplam değişimin % 3.7'si gibi bir oranda kaldığı dolayısıyla modele dahil edilmeyen etkileşimlerin modele dahil edilmesi ile bağımlı değişkeni açıklamada büyük bir farkın sağlanamayacağı görülmektedir. Analizi sonuçlarına göre, (4.1) modeline dahil edilen üç'lü etkileşimlerin bağımlı değişkeni açıklama oranı % 5 gibi düşük bir seviyede kalmıştır.

Tablo-4.5, Tek Sapan Değer Verileri için Ana Etkilere İlişkin Varyans Analizi Sonuçları.

ETKEN		Ortalama	p
Model	AR	0.86278	0.000
	MA	0.89306	
Büyüklik	3 σ	0.83472	0.000
	5 σ	0.92111	
Genişlik	50	0.86417	0.001
	100	0.88500	
	150	0.88458	
Duyarlılık	3.0	0.87250	0.229
	3.5	0.87875	
	4.0	0.88250	
Tür	AO	0.84500	0.000
	IO	0.91083	

Tablo-4.5'te zaman serisi modelleri; AR ve MA, sapan değer büyüklüğü; 3σ ve 5σ , seri genişliği; 50, 100 ve 150, ölçüt değer (C) duyarlılığı; 3.0, 3.5 ve 4.0, sapan değer türü olarak ta AO ve IO biçiminde sınıflandırılmış ve bunlar ortalama değerler cinsinden ele alınmıştır. Söz konusu tablodan modeller arası fark ve sapan değer büyüklüğü açısından farkın önemli olduğu görülmektedir. Seri genişliklerinin 50 ile 100 arasında önemli, 100 ile 150 arasında önemli olmadığı, sapan değer ölçütünün farklı değerleri ile tarama süreçlerinin farklı sonuçlar doğurmadığı, dolayısıyla ölçüt değerinin sapan değeri tespit yönünden önemli olmadığı sonucuna varılmıştır. Sapan değer türü bakımından da farkın önemli olduğu, ulaşılan diğer bir sonuç olarak karşımıza çıkmaktadır. Ortalamalar bakımından iki'li etkileşimleri ayrıntılı bir şekilde görebilmek için aşağıda verilen tablolar incelenebilir.

**Tablo-4.6, Tek Sapan Değer Verileri için
Büyüklik-Model (B*M) Etkileşimi.**

		Model		
		AR	MA	Ortalama
Büyüklik	3σ	0.81389	0.85556	0.834725
	5σ	0.91167	0.93056	0.921115
Ortalama		0.86278	0.89306	0.877920

(Büyüklik x Model) etkileşimini ifade eden Tablo-4.26, gerek model, gerekse büyüklüğün yarattığı farklılığı ifade etmektedir.

**Tablo-4.7, Tek Sapan Değer Verileri için
Tür-Model (T*M) Etkileşimi.**

		Model		
		AR	MA	Ortalama
Tür	AO	0.84222	0.84778	0.845000
	IO	0.88333	0.93833	0.910830
Ortalama		0.862775	0.893055	0.877915

Yukarıdaki tabloda IO'daki AR-MA farkı AO'da gözlenen farktan çok daha büyüktür. Bu sonuç, (Tür x Model) etkileşiminin önemli çıkması ile doğrulanmaktadır.

**Tablo-4.8, Tek Sapan Değer Verileri için
Büyüklik-Tür (B*T) Etkileşimi.**

		Tür		
		AO	IO	Ortalama
Büyüklik	3σ	0.79500	0.89500	0.84500
	5σ	0.87444	0.94722	0.91083
Ortalama		0.83472	0.92111	0.877915

Tablo-4.8 incelendiğinde, Büyüklik olarak tanımlanan 3σ ve 5σ da görülen AO-IO farkı (Büyüklik x Tür) etkileşiminin önemli çıkmasının bir sonucu gibi yorumlanabilir.

**Tablo-4.9, Tek Sapan Değer Verileri için
Tür-Genişlik (T*G) Etkileşimi.**

		Genişlik			Ortalama
		50	100	150	
Tür	AO	0.79917	0.85917	0.87667	0.8450033
	IO	0.92917	0.91083	0.89250	0.9108333
Ortalama		0.86417	0.88500	0.884585	0.8779183

(Tür x Genişlik) etkileşimi için verilen Tablo-4.9 incelendiğinde, 50-100 ölçümleri için elde edilen farklılık göze çarpmaktadır. Tür'e ilişkin elde edilen farklı ortalama değerleri de üzerinde durulması gerekli diğer bir sonuç gibi durmaktadır.

**Tablo-4.10, Tek Sapan Değer Verileri için
Tür-Duyarlılık (T*D) Etkileşimi.**

		Duyarlılık			Ortalama
		C=3.0	C=3.5	C=4.0	
Tür	AO	0.84917	0.84583	0.84000	0.8450000
	IO	0.89583	0.91167	0.92500	0.9108333
Ortalama		0.87250	0.87875	0.88250	0.8779166

(Tür x Duyarlılık) etkileşim tablosu incelendiğinde duyarlılığa ilişkin elde edilmiş ortalamaların oldukça yakın değerler olduğu, farklılığın Tür'den kaynaklandığı görülebilir.

4.2. Çift Sapan Değerlere İlişkin Varyans Analizi Çalışması

Chang Tiao ve Chen (1988) tarafından yapılmış simülasyon sonuçları kullanılarak sapan değer taramada; sapan değer Büyüklüğü (**B**) (3σ , 5σ), Genişlik (**G**) (50, 100, 150), Duyarlılık (**D**) ($C=3.0$, $C=3.5$, $C=4.0$), ve Tür (**T**) (AO, IO) gibi faktörlerin ne derece etkili olduklarını istatistiksel olarak ortaya koymaya yönelik varyans analizi çalışması yapılmıştır. Söz konusu problemlerin çözülmesine yönelik faktörler aşağıdaki gibidir :

Faktör Ana Etkileri;

Sapan Değer Büyüklüğü (B),

Seri Genişliği (G),

Ölçüt Değer Duyarlılığı (D),

Sapan Değer Türleri(T).

Özel seçimli, çok-etkenli ($3 \times 3 \times 2 \times 3$) (2×3^3) deney düzeni için tasarlanan model denklemi aşağıdaki gibidir;

$$Y_{ijkl} = \mu + T_i + D_j + B_k + G_l + (TD)_{ij} + (TB)_{ik} + (TG)_{il} + (DB)_{jk} + (DG)_{jl} + (BG)_{kl} + (TDB)_{ijk} + (TDG)_{ijl} + (TBG)_{ikl} + (DBG)_{jki} + \varepsilon_{ijkl} \quad (4.2)$$

$$i = 1,2; \quad j = 1,2,3; \quad k = 1,2,3; \quad l = 1,2,3.$$

Y_{ijkl} : Simülasyon sonuçları ile elde edilen çift sapan değer tespit olasılıkları.

μ : Genel ortalama.

Tablo-4.11, AR Modelde İki Sapan Değer Tarama Olasılıkları.

		3 σ			5 σ		
		50	100	150	50	100	150
2AO	C=3.0	0.50	0.59	0.63	0.60	0.76	0.80
	C=3.5	0.44	0.54	0.62	0.58	0.76	0.80
	C=4.0	0.38	0.54	0.52	0.53	0.76	0.80
2IO	C=3.0	0.61	0.63	0.65	0.81	0.85	0.82
	C=3.5	0.66	0.65	0.61	0.82	0.85	0.82
	C=4.0	0.58	0.67	0.55	0.82	0.86	0.84
AO+IO	C=3.0	0.65	0.68	0.65	0.86	0.86	0.87
	C=3.5	0.64	0.69	0.65	0.86	0.87	0.88
	C=4.0	0.65	0.63	0.69	0.86	0.87	0.88

Tablo-4.12, (4.2) Modeline İlişkin Tasarım Çizelgesi.

No	Y	T	D	B	G	No	Y	T	D	B	G
1	0.50	1	1	1	1	28	0.82	2	2	1	1
2	0.59	1	1	1	2	29	0.85	2	2	1	2
3	0.63	1	1	1	3	30	0.82	2	2	1	3
4	0.60	1	1	2	1	31	0.58	2	3	2	1
5	0.76	1	1	2	2	32	0.67	2	3	2	2
6	0.80	1	1	2	3	33	0.55	2	3	2	3
7	0.44	1	2	3	1	34	0.82	2	3	3	1
8	0.54	1	2	3	2	35	0.86	2	3	3	2
9	0.62	1	2	3	3	36	0.84	2	3	3	3
10	0.58	1	2	1	1	37	0.65	3	1	1	1
11	0.76	1	2	1	2	38	0.68	3	1	1	2
12	0.80	1	2	1	3	39	0.65	3	1	1	3
13	0.38	1	3	2	1	40	0.86	3	1	2	1
14	0.54	1	3	2	2	41	0.86	3	1	2	2
15	0.52	1	3	2	3	42	0.87	3	1	2	3
16	0.53	1	3	3	1	43	0.64	3	2	3	1
17	0.76	1	3	3	2	44	0.69	3	2	3	2
18	0.80	1	3	3	3	45	0.65	3	2	3	3
19	0.61	2	1	1	1	46	0.86	3	2	1	1
20	0.63	2	1	1	2	47	0.87	3	2	1	2
21	0.65	2	1	1	3	48	0.88	3	2	1	3
22	0.81	2	1	2	1	49	0.65	3	3	2	1
23	0.85	2	1	2	2	50	0.63	3	3	2	2
24	0.82	2	1	2	3	51	0.69	3	3	2	3
25	0.66	2	2	3	1	52	0.86	3	3	3	1
26	0.65	2	2	3	2	53	0.87	3	3	3	2
27	0.61	2	2	3	3	54	0.88	3	3	3	3

Yukarıda verilen tasarım tablosu kullanılarak, çok etkenli varyans analizi işlemi yapılmış ve Tablo-4.13' te verilen gözlem, tahmin ve hata terimleri elde edilmiştir. Aynı analiz işlemi sonucunda, Tablo-4.14' te sunulan sonuçları elde edilmiştir. Tablo-4.12 ile verilen tablonun MINITAB programı ile ANOVA komutunun kullanımı ile yapılmıştır.

- **MINITAB Paket Programı ile Analiz İşlemi :**

```
BATCH
OUTFILE 'SONUC.DAT'
READ C1-C5
0.50 1 1 1 1
0.59 1 1 1 2
0.63 1 1 1 3
0.60 1 1 2 1
0.76 1 1 2 2
0.80 1 1 2 3
0.44 1 2 3 1
0.54 1 2 3 2
0.62 1 2 3 3
0.58 1 2 1 1
.....
.....
.....
0.88 3 3 3 3
END
```

```
NAME C1 'Y'
NAME C2 'T'
NAME C3 'D'
NAME C4 'B'
NAME C5 'G'
```

```
ANOVA Y = T D B G T*D T*B T*G D*B D*G B*G T*D*B T*D*G T*B*G
D*B*G
```

Gözlemlere ilişkin tahmin ve hata değerlerini elde etmek için ise MINITAB'in ANOVA tablosunda ilgili sahaların işaretlenmesi gereklidir. Bu işlemler sonucunda özel seçimli, çok etkenli (3x3x2X3) (2x3³) deney düzeni için sonuçları bulunmuş olur. Analiz sonuçları düzenlenerek Tablo-4.13 ve Tablo-4.14' te verilmiştir.

Tablo-4.13, (4.2) Modeline İlişkin Gözlem, Tahmin ve Hata Değerleri.

No	Gözlem	Tahmin	Hata	No	Gözlem	Tahmin	Hata
1	0.50	0.48	0.02	28	0.82	0.83	-0.01
2	0.59	0.60	-0.01	29	0.85	0.85	0.00
3	0.63	0.44	-0.01	30	0.82	0.82	0.00
4	0.60	0.62	-0.02	31	0.58	0.59	-0.01
5	0.76	0.75	0.01	32	0.67	0.65	0.02
6	0.80	0.79	-0.01	33	0.55	0.56	-0.01
7	0.44	0.43	0.01	34	0.82	0.81	0.01
8	0.54	0.56	-0.02	35	0.86	0.88	-0.02
9	0.62	0.61	0.01	36	0.84	0.83	0.01
10	0.58	0.59	-0.01	37	0.65	0.66	-0.01
11	0.76	0.74	-0.02	38	0.68	0.66	0.02
12	0.80	0.81	-0.01	39	0.65	0.66	-0.01
13	0.38	0.36	0.02	40	0.86	0.85	0.01
14	0.54	0.55	-0.01	41	0.86	0.88	-0.02
15	0.52	0.54	-0.02	42	0.87	0.86	0.01
16	0.53	0.55	-0.03	43	0.64	0.66	-0.02
17	0.76	0.76	0.00	44	0.69	0.67	0.02
18	0.80	0.78	0.02	45	0.65	0.65	0.00
19	0.61	0.62	-0.01	46	0.86	0.85	0.01
20	0.63	0.64	-0.01	47	0.87	0.89	-0.02
21	0.65	0.63	0.02	48	0.88	0.88	0.00
22	0.81	0.80	0.01	49	0.65	0.66	-0.01
23	0.85	0.84	0.01	50	0.63	0.65	-0.02
24	0.82	0.84	-0.02	51	0.69	0.66	0.03
25	0.66	0.66	0.00	52	0.86	0.85	0.01
26	0.65	0.65	0.00	53	0.87	0.85	0.02
27	0.61	0.61	0.00	54	0.88	0.91	-0.03

Tablo-4.14, (4.2) Modeline İlişkin İki Sapan Değer Tarama Verileri için Varyans Analizi Sonuçları.

D.K.	S.D.	K.T.	K.O.	F.	p.
T (Tür)	2	0.202228	0.10111	126.98	0.000
D (Duyarlılık)	2	0.00472	0.00236	2.96	0.090
B (Büyüklik)	1	0.53800	0.53800	675.63	0.000
G (Genişlik)	2	0.05513	0.02757	34.62	0.000
T*D	4	0.00634	0.00159	1.99	0.160
T*B	2	0.00232	0.00116	1.45	0.272
T*G	4	0.07285	0.01821	22.87	0.000
D*B	2	0.00407	0.00203	2.56	0.119
D*G	4	0.00146	0.00037	0.46	0.764
B*G	2	0.00272	0.00136	1.70	0.223
T*D*B	4	0.00190	0.00047	0.60	0.673
T*D*G	8	0.00675	0.00084	1.06	0.448
T*G*B	4	0.00156	0.00039	0.49	0.680
D*B*G	4	0.00080	0.00020	0.25	0.904
Hata	8	0.00800	0.00080		
Toplam	53	0.90883			

Çift sapan değer için yapılan varyans analizi işlemi ve elde edilen varyans analizi tablosu ile sapan değerlere ilişkin Büyüklük (B), seri Genişliği (G), ölçüt değer Duyarlılığı (D) ve sapan değer Türlerinin (T) sapan değerlerin tespitinde ne derece etkili olduğu incelenmiştir.

Analiz İşlemi sonucunda;

Sapan değerlerde Tür (T) (AO, AO), sapan değer Büyüklüğü (B) (3σ , 5σ) ve Seri Genişliklerinin, (50, 100, 150) sapan değerleri tespitinde önemli oldukları, ölçüt değer Duyarlılığının (D) ($c=3.00$, 3.50 , 4.00) önemli olmadığı sonucuna varılmıştır. Elde edilen bu sonuçlar tek sapan değerler analizi için elde edilen sonuçlara paraleldir. (4.2) modeli için planlanan iki'li ve üç'lü etkileşimlerden Tür ve Genişlik etkileşimleri önemli, diğer etkileşimler önemsiz bulunmuştur.

Tablo-4.15, AR Modelde İki Sapan Değer Tarama Verileri için Ana Etkilere İlişkin Varyans Analizi Sonuçları.

ETKEN		Ortalama	p
Büyüklük	3 σ	0.60296	0.000
	5 σ	0.77444	
Genişlik	50	0.61500	0.000
	100	0.72556	
	150	0.72556	
Duyarlılık	3.0	0.66889	0.090
	3.5	0.71000	
	4.0	0.68722	
Tür	2AO	0.61944	0.000
	2IO	0.72778	
	AO+IO	0.71889	

Tablo-4.15'te Ana Etkilere ilişkin sonuçlar incelendiğinde;

- Sapan değerlerin tespit edilmesinde seri uzunluklarının 50 ile 100 arasında önemli, 100'den büyük seriler için sapan değer tespiti yönünden önemli bir değişikliğin olmadığı sonucuna varılmıştır.
- Sapan değerlerin tespitinde ölçüt değer olarak kullanılan C'nin duyarlılığının istatistiksel açıdan önemli olmadığı sonucu görülmektedir.
- Sapan değerlerin tespitinde tek sapan değer olma durumu ve sapan değer büyüklüğünün tespitte etkili olduğu sonucu elde edilmiştir.
- Serilerde çift sapan değer olması durumunda, tarama süreçlerinin güçlerinde bir zayıflama olduğu, dolayısıyla homojen veriler üzerinde bu tip çalışmaların yapılması gerektiği sonucu çıkarılabilir. Ayrıca gözlenmesi gerekli diğer bir sonuç da tarama sürecinin ikinci tip sapan değerleri daha yüksek bir olasılıkla tespit ettiği şeklindedir.

Sapan değerlerin ARIMA modellerle olan ilişkisine bakıldığında, MA modellerde ikinci tip sapan değerlerin daha yüksek bir olasılıkla tespit edildikleri görülebilir. Bu sonuç, MA modellerin hata terimlerinin doğrusal bir bileşeni olarak elde edilmelerinden kaynaklanmaktadır. İkinci tip sapan değerler ise tespit edildikleri pozisyondan sonraki gözlemleri de etkileme özelliğine sahip olduklarından, otokorelasyonla tespite dayalı tarama süreçleri tarafından daha kolay tespit edilmektedir. Birinci tip sapan değerlerin tespit edilmesi ise, söz konusu gözlemden önceki ve sonraki gözlemlerden belirgin bir şekilde farklı olmalarını gerektirir.

4.3. Tarama Süreçlerine İlişkin Varyans Analizi Çalışması

Chang Tiao ve Chen (1988) tarafından yapılmış simülasyon sonuçları kullanılarak; sapan değer taramada; Sapan değer Türü (T) (AO, IO) sapan değer sayısı (Adedi) (A) (1, 2,3) ve sapan değer tarama Süreçleri (S) (LS, M-H, M-B, GM-H, GM-B, Ardışık) gibi faktörlerin ne kadar etkili olduğunu istatistiksel olarak ortaya koymaya yönelik varyans analizi çalışması yapılmıştır.

Faktör Ana Etkileri ;

Sapan Değer Türü (T),

Sapan Değer Sayısı (Adedi) (A),

Tarama Süreçleri (S).

Faktörler için iki' li ve üç'lü etkileşimler; T*A, T*S, A*S, T*A*S.

Özel seçimli, çok-etkenli (2x3x6) deney düzeni için tasarlanmış model denklemi aşağıdaki gibidir :

$$Y_{ijk} = \mu + T_i + A_j + S_k + (TA)_{ij} + (TS)_{ik} + (AS)_{jk} + \varepsilon_{ijk} \quad (4.3)$$

$i = 1,2; \quad j = 1,2,3; \quad k = 1,2,3,4,5,6.$

Y_{ijk} : Tarama süreçlerinin sapan değer tespiti ile meydana gelen hata kareler ortalaması.

μ : Genel ortalama.

(4.3) modeli için hata teriminin tahminlenmesine olanak vermek amacıyla faktörlere ilişkin üç'lü etkileşim modele dahil edilmemiş ve sıfır kabul edilmiştir. Üç'lü etkileşim ile açıklanan değişim hata terimi içerisine dahil edilmiştir.

Tablo-4.16, Tarama Süreçlerine Bağlı Elde Edilen Hata Kareler Ortalamaları.

Sapan Değer	ADET	LS	M-H	M-B	GM-H	GM-B	Ardışık
AO	1	0.0306	0.0185	0.0183	0.0119	0.0100	0.0079*
	2	0.0371	0.0237	0.0245	0.0144	0.0127	0.0089*
	3	0.0665	0.0470	0.0502	0.0242	0.0202	0.0120*
IO	1	0.0071	0.0056	0.0055	0.0055	0.0056	0.0049*
	2	0.0058	0.0046	0.0045	0.0047	0.0054	0.0044*
	3	0.0334	0.0203	0.0217	0.0115	0.0099	0.0052*

Tablo-4.17, (4.3) Modeline İlişkin Tasarım Çizelgesi.

No	Y	T	A	S	No	Y	T	A	S
1	0.0306	1	1	1	19	0.0071	2	1	1
2	0.0185	1	1	2	20	0.0056	2	1	2
3	0.0183	1	1	3	21	0.0055	2	1	3
4	0.0119	1	1	4	22	0.0055	2	1	4
5	0.0100	1	1	5	23	0.0056	2	1	5
6	0.0079	1	1	6	24	0.0049	2	1	6
7	0.0371	1	2	1	25	0.0058	2	2	1
8	0.0237	1	2	2	26	0.0046	2	2	2
9	0.0245	1	2	3	27	0.0045	2	2	3
10	0.0144	1	2	4	28	0.0047	2	2	4
11	0.0127	1	2	5	29	0.0054	2	2	5
12	0.0089	1	2	6	30	0.0044	2	2	6
13	0.0665	1	3	1	31	0.0334	2	3	1
14	0.0470	1	3	2	32	0.0203	2	3	2
15	0.0502	1	3	3	33	0.0217	2	3	3
16	0.0242	1	3	4	34	0.0115	2	3	4
17	0.0202	1	3	5	35	0.0099	2	3	5
18	0.0120	1	3	6	36	0.0052	2	3	6

Tablo-4.17'de verilen çizelge kullanılarak, MINITAB paket programı ile çok etkenli varyans analizi yöntemi yapılmış ve Tablo-4.13' te verilen gözlem, tahmin ve hata değerleri elde edilmiştir. Aynı analiz işlemi ile, Tablo-4.14' te verilen varyans analizi sonuçları elde edilmiştir. Tablo-4.12 ile verilen tablonun MINITAB paket programı ile analiz edilme işlemi aşağıda tanımlandığı şekliyle ANOVA komutunun kullanımı ile yapılmıştır.

- **MINITAB Paket Programı ile Analiz İşlemi :**

```
BATCH
OUTFILE 'SONUC.DAT'
READ C1-C5
0.0306 1 1 1
0.0185 1 1 2
0.0183 1 1 3
0.0119 1 1 4
0.0100 1 1 5
0.0079 1 1 6
0.0371 1 2 1
0.0237 1 2 2
0.0245 1 2 3
0.0144 1 2 4
.....
.....
.....
0.0052 2 3 6
END
```

```
NAME C1 'Y'
NAME C2 'T'
NAME C3 'A'
NAME C4 'S'
```

ANOVA Y = T A S T*A T*S A*S

Gözlemlere ilişkin tahmin ve hata değerlerini elde etmek için ise MINITAB'in ANOVA tablosunda ilgili sahalardan işaretlenmesi gereklidir. Bu işlemler sonucunda özel seçimli, çok-etkenli (2x3x6) deney düzeni için analiz sonuçları elde edilmiş olur. Elde edilen analiz sonuçları, düzenlenerek Tablo-4.18 ve Tablo-4.19' da verilmiştir.

Tablo-4.18, (4.3) Modeline İlişkin Gözlem Tahmin ve Hata Değerleri.

No	Gözlem	Tahmin	Hata	No	Gözlem	Tahmin	Hata
1	0.0306	0.0312	-0.0006	19	0.0071	0.0065	0.0006
2	0.0185	0.0195	-0.0010	20	0.0056	0.0046	0.0010
3	0.0183	0.0198	-0.0015	21	0.0055	0.0040	0.0015
4	0.0119	0.0111	0.0008	22	0.0055	0.0062	-0.0007
5	0.0100	0.0091	0.0009	23	0.0056	0.0065	-0.0009
6	0.0079	0.0065	0.0014	24	0.0049	0.0064	-0.0015
7	0.0371	0.0362	0.0009	25	0.0058	0.0067	-0.0009
8	0.0237	0.0240	-0.0003	26	0.0046	0.0043	0.0003
9	0.0245	0.0248	-0.0003	27	0.0045	0.0042	0.0003
10	0.0144	0.0144	0.0000	28	0.0047	0.0047	0.0000
11	0.0127	0.0128	-0.0001	29	0.0054	0.0053	0.0001
12	0.0089	0.0091	-0.0002	30	0.0044	0.0042	0.0002
13	0.0665	0.0668	-0.0003	31	0.0334	0.0330	0.0004
14	0.0470	0.0457	0.0013	32	0.0203	0.0216	-0.0013
15	0.0502	0.0484	0.0018	33	0.0217	0.0235	-0.0018
16	0.0242	0.0249	-0.0007	34	0.0115	0.0108	0.0007
17	0.0202	0.0210	-0.0008	35	0.0099	0.0091	0.0008
18	0.0120	0.0132	-0.0012	36	0.0052	0.0040	0.0012

Tablo-4.19, Tarama Süreçlerine Bağlı Elde Edilen Hata Kareler Ortalaması Verilerine İlişkin Varyans Analizi Sonuçları.

D.K.	S.D.	K.T.	K.O.	F	P
T (Tür)	1	0.002070	0.002070	662.02	0.000
A(Adet)	2	0.001837	0.000918	293.65	0.000
S(Süreç)	5	0.002129	0.000426	136.15	0.000
T*A	2	0.000127	0.000063	20.24	0.000
T*S	5	0.000671	0.000134	42.92	0.000
A*S	10	0.000787	0.000079	25.16	0.000
Hata	10	0.000031	0.000003		
Toplam	35	0.007652			

Eldeki veri seti için yapılan varyans analizi işlemi ve elde edilen varyans analizi tablosu ile sapan değerlerin Tür (T), sapan değer sayıları (A) ve tarama süreçleri (S) arasındaki ilişki ele alınmış, etkileri incelenmiştir. Sözü edilen üç faktör için iki'li etkileşimler de incelenerek sonuçlar elde edilmiştir. Tablo-4.19'da verilen varyans analizi faktör ana etkilerine ilişkin sonuçlar incelendiğinde;

Tablo-4.20, Tarama Süreçleri ve Diğer Faktör Ana Etkilerine İlişkin Varyans Analizi Sonuçları.

ETKEN		Ortalama	p
Tür	AO	0.0244	0.000
	IO	0.0092	
Adet	1	0.0109	0.000
	2	0.0125	
	3	0.0268	
Süreç	LS	0.0300	0.000
	M-H	0.0199	
	M-B	0.0208	
	GM-H	0.0120	
	GM-B	0.0106	
	Ardışık	0.0072	

- Yukarıdaki tablo'da tanımlı hata kareler ortalaması sonuçlarına bakıldığında sapan değer Türü'nün (AO, IO) (T) büyük oranda önemli bir faktör olduğu görülmektedir.
- İkinci faktör olarak belirlenen sapan değer sayısının (1,2,3) (A) önemli bulunduğu, dolayısıyla seri içerisinde sapan değer sayısı arttıkça tarama süreçlerinin sapan değer tespit gücü önemli oranda zayıflamaktadır.
- Üçüncü faktör, tarama süreçlerinin performansı ile ilgili olanıdır. Tarama süreçlerine ilişkin performans analiz edildiğinde, süreçler arası farkın istatistik olarak önemli bulunduğu görülmektedir. Bu üç faktör arasında oluşturulan iki'li etkileşimler de önemli bulunmuştur. Etkileşimler için ayrıntılı dökümler, hata kareler ortalaması bazında aşağıda Tablo-4.21, Tablo-4.22 ve Tablo-4.23'te ifade edilmiştir.

**Tablo-4.21, Hata Kareler Ortalaması Verileri için
Tür x Adet (T*A) Etkileşimi.**

		Adet			Ortalama
		1	2	3	
Tür	AO	0.0162	0.0202	0.0367	0.0244
	IO	0.0057	0.0049	0.0170	0.0092
	Ortalama	0.0109	0.0125	0.0268	0.0168

(Tür x Adet) etkileşim tablosundan görülebileceği gibi sapan değer türleri (AO, IO) arasında ortalama bazında meydana gelen farklılık, sapan değer sayısının artmasıyla belirginleşmiş Bunun sonucunda iki'li etkileşim önemli bulunmuştur.

**Tablo-4.22, Hata Kareler Ortalaması Verileri için
Tür x Süreç (T*S) Etkileşimi.**

		Süreç						Ortalama
		LS	M-H	M-B	GM-H	GM-B	ARD	
Tür	AO	0.0447	0.0297	0.0310	0.0168	0.0143	0.0096	0.0244
	IO	0.0154	0.0101	0.0106	0.0072	0.0069	0.0048	0.0092
	Ortalama	0.0300	0.0199	0.0208	0.0120	0.0106	0.0072	0.0168

(Tür x Süreç) etkileşim tablosu incelendiğinde hata kareler ortalaması bazında minimum değerler, ardışık yöntemin bulunduğu sütunda gözlenmektedir. Sapan değer türlerinden AO sapan değerinin ortaya çıkması hata kareler ortalamasında ciddi artışlara sebep olmaktadır.

**Tablo-4.23, Hata Kareler Ortalaması Verileri için
Adet x Süreç (A*S) Etkileşimi.**

		Süreç						
		LS	M-H	M-B	GM-H	GM-B	ARD	Ortalama
Adet	1	0.0188	0.0120	0.0119	0.0087	0.0078	0.0064	0.0109
	2	0.0214	0.0141	0.0145	0.0095	0.0090	0.0066	0.0125
	3	0.0499	0.0336	0.0359	0.0178	0.0150	0.0086	0.0268
	Ortalama	0.0300	0.0199	0.0208	0.0120	0.0106	0.0072	0.0168

(Adet x Süreç) etkileşim tablosu, sapan değer sayısının artışı ile hata kareler ortalamasının arttığını ifade etmektedir.

4.4. Uygulama

Teorik yapısı ve özellikleri tanımlı sapan değerlerin tespit edilmesi amacına yönelik üç farklı veri seti üzerinde uygulama çalışması planlanmıştır.

Birinci tip veriler, bir çok çalışmada kullanılan, Box-Jenkins A Serisi veriler olarak bilinen ve iki saat'te bir elde edilmiş kimyasal konsantrasyon ölçümlerini ifade etmektedir. Söz konusu veri seti, birinci ve ikinci tip sapan değerleri içerdiğinden, tanımlayıcı özelliği dolayısıyla kullanılmıştır (*Box-jenkins, 1976*).

İkinci tip veriler, ÇİMENTAŞ T.A.Ş 1980-1987 aylık çimento ihracat verileri olup, düzeltme amaçlı veriler olarak düşünülmüştür. Temel analiz gereği kullanılan veriler üzerinde sapan değer analizi yapılmıştır (*Çimentaş T.A.Ş, 1995*).

Üçüncü tip veriler, gıda sektöründe bir un fabrikasında unların su tutma kapasitesini belirlemek amacıyla destile su ile un örneklerinin karıştırılıp santrifüj edilmesi sonucu tutulan suyun ölçülmesi esasına dayanan bir yöntemdir. Söz konusu veriler elde edilirken, yüksek devire dayanıklı plastik deney tüplerinin üzerine 5 ml destile su ilave edilir. Daha sonra 25 kez çalkalanır. Bu tip işlem her 5 dakikada bir tekrar edilir. 20 dakika sonunda devri 5000 olan santrifüjde 15 dakika santrifüjlenir ve süzüntü ayrılır. Tüpler birkaç kez kuvvetlice silkelenerek 45 °C eğimli yerde 10 dakika bekletildikten sonra tartılır. Aradaki farktan % su tutma kapasitesi hesaplanır (*Doğan S., Ünal S.S., 1990*).

Farklı amaçlarla analizi yapılan yukarıda tanımlı bu üç tür veri, zaman serileri model aşamalarında MINITAB istatistik paket programı ile, sapan değer analizi aşamalarında; diğer yöntemlere göre etkinliği varyans analizi yöntemi ile belirlenen, Ardışık Yöntem ile yapılmıştır. Ardışık Yöntem ile analiz yapılırken ölçüt değer duyarlılığı C=3.00 alınmış ve yüksek duyarlılıkla analiz yapılmıştır. Sapan değer analizi yapılırken; verilerin uyum gösterdiği zaman serisi modelleri elde edilmiş, elde edilen parametre değerleri, ölçüt değer duyarlılığı, gözlem ve hata değerleri kullanılarak Ardışık Yöntem ile sapan değer tarama işlemleri yapılmıştır.

4.1.1. Kontrol Amaçlı

Sapan değerler analizi, üretim faktörlerinin bir veya bir kaçında zaman zaman meydana gelen aksaklıkları belirlemek için kullanılmaktadır. Bu amaçla Box-Jenkins A serisi olarak bilinen veriler üzerinde sapan değer analizi yapılarak ortaya çıkan sonuçlar üzerinde gerekli yorumlar yapılmıştır.

Tablo-4.24, Kontrol Amaçlı Box-Jenkins A Serisi Gözlem Değerleri (197 Adet).

No	Göz.	No	Göz.	No	Göz.	No	Göz.	No	Göz.	No	Göz.
1	17.0	36	17.1	71	17.3	106	16.7	141	16.6	176	17.0
2	16.6	37	17.6	72	17.4	107	16.2	142	16.5	177	16.9
3	16.3	38	17.7	73	17.7	108	16.6	143	17.0	178	17.1
4	16.1	39	17.4	74	16.8	109	16.9	144	16.7	179	17.2
5	17.1	40	17.8	75	16.9	110	16.5	145	16.7	180	17.4
6	16.9	41	17.6	76	17.0	111	16.6	146	16.9	181	17.5
7	16.8	42	17.5	77	16.9	112	16.6	147	17.4	182	17.9
8	17.4	43	16.5*	78	17.0	113	17.0	148	17.1	183	17.0
9	17.1	44	17.8	79	16.6	114	17.1	149	17.0	184	17.0
10	17.0	45	17.3	80	16.7	115	17.1	150	16.8	185	17.0
11	16.7	46	17.3	81	16.8	116	16.7	151	17.2	186	17.2
12	17.4	47	17.1	82	16.7	117	16.8	152	17.2	187	17.3
13	17.2	48	17.4	83	16.4	118	16.3	153	17.4	188	17.4
14	17.4	49	16.9	84	16.5	119	16.6	154	17.2	189	17.4
15	17.4	50	17.3	85	16.4	120	16.8	155	16.9	190	17.0
16	17.0	51	17.6	86	16.6	121	16.9	156	16.8	191	18.0
17	17.3	52	16.9	87	16.5	122	17.1	157	17.0	192	18.2
18	17.2	53	16.7	88	16.7	123	16.8	158	17.4	193	17.6
19	17.4	54	16.8	89	16.4	124	17.0	159	17.2	194	17.8
20	16.8	55	16.8	90	16.4	125	17.2	160	17.2	195	17.7
21	17.1	56	17.2	91	16.2	126	17.3	161	17.1	196	17.2
22	17.4	57	16.8	92	16.4	127	17.2	162	17.1	197	17.4
23	17.4	58	17.6	93	16.3	128	17.3	163	17.1		
24	17.5	59	17.2	94	16.4	129	17.2	164	17.4		
25	17.4	60	16.6	95	17.0	130	17.2	165	17.2		
26	17.6	61	17.1	96	16.9	131	17.5	166	16.9		
27	17.4	62	16.9	97	17.1	132	16.9	167	16.9		
28	17.3	63	16.6	98	17.1	133	16.9	168	17.0		
29	17.0	64	18.0*	99	16.7	134	16.9	169	16.7		
30	17.8	65	17.2	100	16.9	135	17.0	170	16.9		
31	17.5	66	17.3	101	16.5	136	16.5	171	17.3		
32	18.1	67	17.0	102	17.2	137	16.7	172	17.8		
33	17.5	68	16.9	103	16.4	138	16.8	173	17.8		
34	17.4	69	17.3	104	17.0	139	16.7	174	17.6		
35	17.4	70	16.8	105	17.0	140	16.7	175	17.5		

Tablo-4.25, Box-Jenkins A Serisi için Değişik İterasyonlardaki Parametre Tahminleri.

Parametre Tahminleri				
İterasyon	HKT	AR	MA	Sabit
0	33.2117	0.100	0.100	15.446
1	24.2840	0.250	-0.050	12.861
2	22.9740	0.400	0.080	10.287
3	21.6090	0.550	0.210	7.713
4	20.2822	0.700	0.338	5.141
5	19.3644	0.850	0.467	2.570
6	19.2640	0.891	0.525	1.869
7	19.2467	0.903	0.554	1.659
8	19.2422	0.909	0.568	1.554
9	19.2409	0.912	0.575	1.501
10	19.2406	0.914	0.579	1.474
11	19.2405	0.914	0.581	1.460
12	19.2405	0.915	0.582	1.453
13	19.2405	0.915	0.583	1.449

Zaman serilerinde model belirleme ve parametre tahminleme aşamaları ikinci bölümün (2.4) kısmında "Zaman Serilerinin Analizi" başlığı altında adımsal olarak verilmiştir. Modellerin yeterliliği her uygulamanın "Parametre Tahmin Değerleri" tablosunda *Box-Pierce* istatistiği ile de doğrulanmıştır. Bunun için parametrelere ilişkin t değerlerine veya ilgili tabloların koyu zemin ile ifade edilen bölümlerine bakılabilir.

Tablo-4.26, Box-Jenkins A Serisi için Parametre Tahmin Değerleri.

Parametre Tahmin Değerleri			
Parametre	Tahmini	St. Sapma	t-değeri
AR(1)	0.9151	0.0433	21.11
MA(1)	0.5828	0.0849	6.87
Sabit	1.4490	0.0094	154.36
Ortalama	17.0656	0.1105	
HKT	19.188		
HKO	0.0989		
Box-Pierce	24 Lag	28.00	
Tablo Değ.	22 s.d	33.92	
Tablo değ > Box-Pierce old. model uygun.			

Tablo-4.27, Box-Jenkins A Serisi Gözlem, Tahmin ve Hata Değerleri.

No	Gözlem	Tahmin	Hata	No	Gözlem	Tahmin	Hata
1	17.0	16.8767	0.12325	48	17.4	17.2022	0.19781
2	16.6	16.9337	-0.33374	49	16.9	17.2563	-0.35633
3	16.3	16.8340	-0.53403	50	17.3	17.1217	0.17828
4	16.1	16.6762	-0.57623	51	17.6	17.1762	0.42380
5	17.1	16.5178	0.58219	52	16.9	17.3076	-0.40764
6	16.9	16.7578	0.14222	53	16.7	17.1516	-0.45163
7	16.8	16.8312	-0.03118	54	16.8	16.9942	-0.19424
8	17.4	16.8407	0.55928	55	16.8	16.9358	-0.13575
9	17.1	17.0457	0.05433	56	17.2	16.9017	0.29833
10	17.0	17.0654	-0.06541	57	16.8	16.9017	-0.21472
11	16.7	17.0437	-0.34369	58	17.6	16.9477	0.65231
12	17.4	16.9313	0.46866	59	17.2	17.1745	0.02553
13	17.2	17.0985	0.10152	60	16.6	17.1737	-0.57371
14	17.4	17.1294	0.27058	61	17.1	16.9739	0.12612
15	17.4	17.2139	0.18608	62	16.9	17.0236	-0.12358
16	17.0	17.2632	-0.26316	63	16.6	16.9861	-0.38608
17	17.3	17.1589	0.14106	64	18.0*	16.8645	1.13547
18	17.2	17.1979	0.00211	65	17.2	17.2589	-0.05893
19	17.4	17.1874	0.21264	66	17.3	17.2229	-0.07707
20	16.8	17.2477	-0.44768	67	17.0	17.2352	-0.23518
21	17.1	17.0835	0.01655	68	16.9	17.1426	-0.24623
22	17.4	17.0874	0.31256	69	17.3	17.0555	0.24454
23	17.4	17.1894	0.21055	70	16.8	17.1376	-0.33758
24	17.5	17.2489	0.25110	71	17.3	17.0193	0.28071
25	17.4	17.3168	0.08322	72	17.4	17.1165	0.28350
26	17.6	17.3231	0.27690	73	17.7	17.2064	0.49361
27	17.4	17.3933	0.00675	74	16.8	17.3585	-0.55846
28	17.3	17.3677	-0.06768	75	16.9	17.1480	-0.24802
29	17.0	17.3195	0.31954	76	17.0	17.0586	-0.05860
30	17.8	17.1918	0.60821	77	16.9	17.0397	-0.13972
31	17.5	17.3832	0.11681	78	17.0	16.9955	0.00451
32	18.1	17.3950	0.70496	79	16.6	17.0029	-0.40294
33	17.5	17.6013	-0.10133	80	16.7	16.8744	-0.17436
34	17.4	17.5222	-0.12217	81	16.8	16.8327	-0.03266
35	17.4	17.4428	-0.04280	82	16.7	16.8416	-0.14158
36	17.1	17.3966	-0.29655	83	16.4	17.8136	0.41355
37	17.6	17.2699	0.33010	84	16.5	16.6975	0.19753
38	17.7	17.3622	0.33775	85	16.4	16.6631	-0.26314
39	17.4	17.4493	-0.04930	86	16.6	16.6099	-0.00987
40	17.8	17.4003	0.39966	87	16.5	16.6453	-0.14528
41	17.6	17.5047	0.09528	88	16.7	16.6327	0.06731
42	17.5	17.4991	0.00090	89	16.4	16.6918	-0.29182
43	16.5*	17.4626	-0.96259	90	16.4	16.6266	-0.22658
44	17.8	17.1090	0.69099	91	16.2	16.5886	-0.38856
45	17.3	17.3349	-0.03494	92	16.4	16.4999	-0.09994
46	17.3	17.3005	-0.00465	93	16.3	16.5148	-0.21476
47	17.1	17.2804	-0.18036	94	16.4	16.4902	-0.09016

Tablo-4.28, Box-Jenkins A Serisi Gözlem, Tahmin ve Hata Değerleri Devamı.

No	Gözlem	Tahmin	Hata	No	Gözlem	Tahmin	Hata
95	17.0	16.5091	0.49094	141	16.6	16.7936	-0.19362
96	16.9	16.7195	0.18054	142	16.5	16.7524	-0.25237
97	17.1	16.8088	0.29116	143	17.0	16.6951	0.30490
98	17.1	16.9274	0.17260	144	16.7	16.8279	-0.12788
99	16.7	16.9965	-0.29649	145	16.7	16.8056	-0.10557
100	16.9	16.9038	-0.00383	146	16.9	16.7926	0.10743
101	16.5	16.9163	-0.41629	147	17.4	16.8514	0.54855
102	17.2	16.7906	0.40937	148	17.1	17.0519	0.04808
103	16.4	16.9500	-0.55001	149	17.0	17.0691	-0.06906
104	17.0	16.7771	0.22295	150	16.8	17.0458	-0.24582
105	17.0	16.8756	0.12436	151	17.2	16.9658	0.23419
106	16.7	16.9331	-0.23309	152	17.2	17.0521	0.14790
107	16.2	16.8669	-0.66689	153	17.4	17.1024	0.29760
108	16.6	16.6621	-0.06215	154	17.2	17.1982	0.00183
109	16.9	16.6758	0.22425	155	16.9	17.1875	-0.28752
110	16.5	16.7834	-0.28337	156	16.8	17.0816	-0.28162
111	16.6	16.7132	-0.11317	157	17.0	16.9867	0.01332
112	16.6	16.7055	-0.10549	158	17.4	16.9978	0.40220
113	17.0	16.7010	0.21264	159	17.2	17.1372	0.06279
114	17.1	16.8313	0.26858	160	17.2	17.1520	0.04800
115	17.1	16.9405	0.15950	161	17.1	17.1606	-0.06061
116	16.7	17.0041	-0.30412	162	17.1	17.1324	-0.03240
117	16.8	16.9083	-0.10828	163	17.1	17.1160	-0.01596
118	16.3	16.8857	-0.58566	164	17.4	17.1064	0.29362
119	16.6	16.7063	-0.10632	165	17.2	17.2005	-0.00049
120	16.8	16.7015	0.09851	166	16.9	17.1889	-0.28887
121	16.9	16.7651	0.13486	167	16.9	17.0824	-0.18241
122	17.1	16.8355	0.26453	168	17.0	17.0204	-0.02037
123	16.8	16.9429	-0.14291	169	16.7	17.0174	-0.31744
124	17.0	16.9058	0.09416	170	16.9	16.9160	-0.01604
125	17.2	16.9507	0.24931	171	17.3	16.9234	0.37659
126	17.3	17.0433	0.25670	172	17.8	17.0606	0.73937
127	17.2	17.1305	0.06951	173	17.8	17.3067	-0.49325
128	17.3	17.1481	0.15192	174	17.6	17.4502	0.14982
129	17.2	17.1916	0.00844	175	17.5	17.4673	0.03269
130	17.2	17.1837	0.01633	176	17.0	17.4441	-0.44407
131	17.5	17.1791	0.32093	177	16.9	17.2644	-0.36437
132	16.9	17.2761	-0.37608	178	17.1	17.1264	-0.02641
133	16.9	17.1332	-0.23324	179	17.2	17.1125	0.08753
134	16.9	17.0500	-0.14999	180	17.4	17.1376	0.26242
135	17.0	17.0015	-0.00147	181	17.5	17.2187	0.28133
136	16.5	17.0064	-0.50643	182	17.9	17.2992	0.60084
137	16.7	16.8432	-0.14316	183	17.0	17.4790	-0.47899
138	16.8	16.8145	-0.01448	184	17.0	17.2847	-0.28472
139	16.7	16.8310	-0.13099	185	17.0	17.1715	-0.17150
140	16.7	16.8074	0.10738	186	17.2	17.1055	0.09448

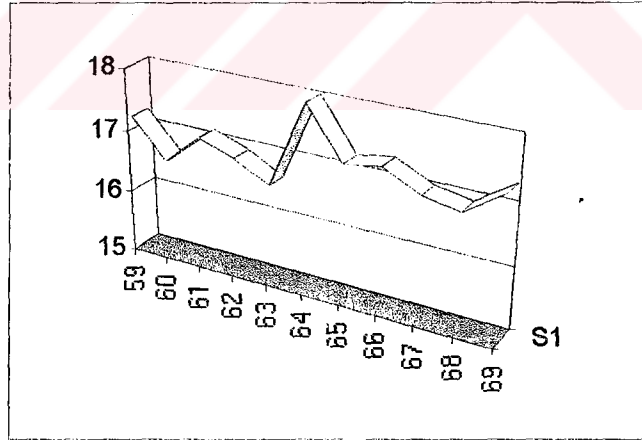
Tablo-4.29, Box-Jenkins A Serisi Gözlem, Tahmin ve Hata Değerleri Devamı.

No	Gözlem	Tahmin	Hata	No	Gözlem	Tahmin	Hata
187	17.3	17.1335	0.16647	193	17.6	17.6588	-0.05883
188	17.4	17.1831	0.21692	194	17.8	17.5889	0.21109
189	17.4	17.2452	0.15481	195	17.7	17.6146	0.08538
190	17.0	17.2814	-0.28138	196	17.2	17.5964	-0.39638
191	18.0	17.1696	0.83044	197	17.4	17.4196	-0.01959
192	18.2	17.4367	0.76331				

Tablo-4.30, Box-Jenkins A serisi Kuşkuğu Gözlem Serisi(59-69. Gözlemler).

1	2	3	4	5	Kuşkuğu	5	4	3	2	1
59	60	61	62	63	64	65	66	67	68	69
17.2	16.6	17.1	16.9	16.6	18.0	17.2	17.3	17.0	16.9	17.3

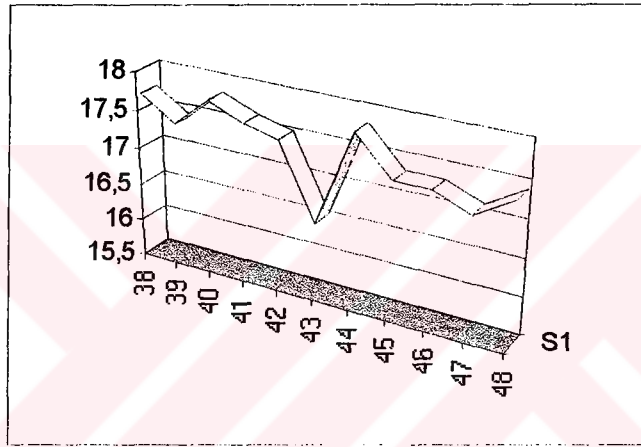
Grafik-4.1, 64. Pozisyon için Kuşkuğu Gözlem Serisi.



Tablo-4.31, Box-Jenkins A serisi Kuşkulu Gözlem Serisi(38-48. Gözlemler).

1	2	3	4	5	Kuşkulu	5	4	3	2	1
38	39	40	41	42	43	44	45	46	47	48
17.7	17.4	17.8	17.6	17.5	16.5	17.8	17.3	17.3	17.1	17.4

Grafik-4.2, 43. Pozisyon için Kuşkulu Gözlem Serisi.



Tablo-4.32, Box-Jenkins A Serisi için Sapan Değer Analizi Sonuçları.

Sapan Değer Analizi Sonuçları-1				
Parametreler		Sapan Değer		Döngü
AR	MA	AO	IO	1
0.91	0.58	43.Pozisyon	64.Pozisyon	

Kontrol amaçlı verilere sapan değer analizi uygulandığında, verilerin 43 ve 64'üncü pozisyonlarında sapan değere rastlandığı görülmektedir. 43'üncü pozisyonda meydana gelen birinci tip sapan değer; sisteme dayanmayan, kişilerden, makinalardan veya makinaların hatalı kullanımlarından doğan bir sonuçtur. 64'üncü pozisyonda meydana gelen hata ise, muhtemelen bozuk hammadde veya makinalarda yapılması zorunlu bir ayarlamayı gerekli kılmaktadır. 43 ve 64'üncü gözlem değerleri yerine, sapan

olan etki (3.10) ve (3.11) eşitlikleri ile tahminlenerek veriler değiştirilmekte, değiştirilmiş gözlem değerlerinin seriye dahil edilmesi ile yeniden sapan değer analizi yapılmaktadır. Söz konusu değişim işleminden sonra, elde edilen sonuçlar Tablo-4.33'te ifade edilmiştir.

Tablo-4.33, Box-Jenkins A Serisi için Sapan Değer Analizi Sonuçları.

Sapan Değer Analizi Sonuçları-2				
Sapan Değer Etkileri Giderildikten Sonra				
AR	MA	AO	IO	
0.89	0.47	Rastlanmadı	Rastlanmadı	2

Değiştirilmiş veri değerleri seriye dahil edildikten sonra seri içerisinde sapan değere rastlanmamıştır. Bu arada 0.91 olarak elde edilmiş AR parametre değeri, 0.89 olarak 0.58 olarak elde edilmiş MA parametre değerinin 0.47 olarak değişmiştir. Burada MA parametresinde meydana gelen değişimin daha büyük olduğu görülmektedir.

4.4.2. Düzeltme Amaçlı (Temel Analiz Gereği)

Sapan değer analizi, yapılması planlanan analiz türü ne olursa olsun temel analiz özelliği taşımaktadır. Heterojen yapı arz eden veri setleri için bir veya birden çok veri seti üzerinde düzeltme işlevi görür. Bu sayede yapılması planlanan analiz sonucunda elde edilecek parametre değerlerinin yanlış (underestimate veya overestimate) olmasına neden olan, muhtemel şartların bir veya bir kaç ı ortadan kalkmış olacaktır. Bu amaçla yukarıdaki veri setinden farklı olarak, İzmir ÇİMENTAŞ T. A. Ş'nin 1980-1987 tarihlerinde aylık bazda ihracat verileri kullanılmıştır.

**Tablo-4.34, İzmir ÇİMENTAŞ T.A.Ş'nin
1980-1987 İhracat Rakamları (1000 ton).**

	YILLAR							
	1980	1981	1982	1983	1984	1985	1986	1987
Ocak	50.102	42.225	77.201	43.706	49.503	70.711	59.683	72.633
Şubat	71.145	40.914	68.924	46.320	67.434	52.727	66.256	86.870
Mart	70.501	55.103	64.842	77.115	68.741	86.862	89.786	79.721
Nisan	83.520	77.122	61.873	65.662	69.010	81.940	101.769	103.820
Mayıs	69.563	67.886	66.160	106.352	72.785	86.384	100.492	91.222
Haziran	72.458	71.705	84.226	52.389	57.894	63.961	79.214	91.223
Temmuz	60.688	67.705	87.153	74.530	60.025	80.405	97.618	83.663
Agustos	55.644	82.760	85.334	101.974	86.433	85.255	77.594	76.528
Eylül	76.358	91.458	64.612	66.265	78.107	87.404	98.737	84.026
Ekim	62.741	84.560	70.985	53.068	91.288	83.804	104.815	100.705
Kasım	76.588	60.839	67.395	73.906	78.373	79.582	83.386	83.752
Aralık	54.032	68.323	55.755	41.513	68.427	77.393	73.309	84.062

**Tablo-4.35, ÇİMENTAŞ T.A.Ş verileri için
Değişik İterasyonlardaki Parametre Tahminleri.**

Parametre Tahmin Tahminleri			
İterasyon	HKT	AR	Sabit
0	20304.8	0.100	66.903
1	18620.2	0.250	55.740
2	17892.7	0.400	44.557
3	17860.4	0.435	41.935
4	17860.0	0.438	41.631
5	17860.0	0.439	41.595
6	17860.0	0.439	41.591

**Tablo-4.36, ÇİMENTAŞ T.A.Ş verileri için
Parametre Tahmin Değerleri.**

Parametre Tahmin Değerleri			
Parametre	Tahmini	St. Sapma	t-değer
AR(1)	0.4389	0.0928	4.73
Sabit	41.591	1.403	29.64
Ortalama	74.123	2.501	
HKT	17770		
HKO	189.0		
Box-Pierce	24 Lag	26.50	
Tablo Deg.	23 s.d	35.17	
Tablo değ > Box-Pierce old. model uygun.			

**Tablo-4.37, ÇİMENTAŞ T.A.Ş Verileri için
Sapan Değer Analizi Sonuçları.**

Sapan Değer Analizi Sonuçları-1		
Parametreler	Sapan Değer	Döngü
AR 0.4389	AO 41. Pozisyon	1

ÇİMENTAŞ ihracat verilerine uygulanan sapan değer analiz işlemi sonucunda 41'inci pozisyonda sapan değere rastlanmıştır. 41'inci pozisyona karşılık gelen 1983 yılı mayıs ayında meydana gelen 106.352 bin ton'luk ihracat rakamı beklenenden çok büyük (28.174 bin ton daha büyük) bir rakam olduğundan sapan değer olarak elde edilmiştir. (3.10) eşitliği ile hesap edilen 28.174 büyüklüğündeki etki değerinin çıkarılması elde edilen düzeltilmiş verinin seriye dahil edilmesi yeniden sapan değer analizi yapılmıştır. Yapılan analiz işlemi sonucuna elde sonuçlar Tablo-4.38'te verilmiştir.

**Tablo-4.38, ÇİMENTAŞ T.A.Ş Verileri için
Sapan Değer Analizi Sonuçları.**

Sapan Değer Analizi Sonuçları-2		
Sapan Değer Etkileri Giderildikten Sonra		
Parametreler	Sapan Değer	Döngü
AR 0.503	AO Rastlanmadı	2

Düzeltilmiş verilere uygulanan sapan değer analizi sonucunda sapan değere rastlanmamıştır.

4.4.3 Etkinlik Artırma Amaçlı

Etkinlik artırma amacına yönelik olarak kullanılan gıda sektöründe unların su tutma yüzdelere ilişkin veriler üzerinde yapılacak analiz işlemi sonucunda varyansın tahmini olan hata kareler ortalaması ve benzer göstergelerde meydana gelecek değişimler incelenecektir. Bu amaçla benzer yöntem ve eşit zaman aralıklarında elde edilen 72 adet veri kullanılmaktadır.

Tablo-4.39, Gıda Sektöründe Unların Su Tutma Yüzde Değerleri.

No	Göz.	No	Göz.	No	Göz.	No	Göz.	No	Göz.	No	Göz.
1	55.70	13	58.67	25	55.25	37	59.12	49	56.65	61	58.84
2	55.08	14	59.83	26	56.27	38	58.75	50	56.59	62	59.08
3	55.02	15	55.65	27	53.58	39	62.56	51	56.07	63	58.50
4	55.01	16	57.19	28	56.17	40	62.28	52	56.99	64	59.05
5	56.15	17	60.78	29	56.19	41	62.93	53	58.88	65	58.50
6	56.38	18	60.55	30	56.24	42	62.64	54	60.83	66	59.78
7	56.20	19	56.97	31	55.83	43	55.99	55	58.57	67	60.96
8	56.82	20	56.17	32	56.51	44	56.58	56	61.23	68	60.55
9	56.64	21	57.16	33	54.38	45	56.18	57	57.58	69	60.36
10	56.81	22	56.28	34	55.15	46	56.35	58	61.75	70	61.26
11	56.18	23	63.17	35	60.33	47	56.88	59	58.25	71	61.08
12	56.95	24	58.79	36	61.04	48	56.50	60	58.95	72	61.04

Tablo-4.40, Gıda Sektöründe Unların Su Tutma Yüzde Değerleri için Değişik İterasyonlardaki Parametre Değerleri.

Parametre Tahmin Değerleri			
İterasyon	HKT	AR	Sabit
0	353.094	0.100	52.330
1	304.022	0.250	43.604
2	272.113	0.400	34.876
3	257.370	0.550	26.147
4	256.268	0.598	23.326
5	256.256	0.603	23.066
6	256.256	0.603	23.038

Tablo-4.41, Gıda Sektöründe Unların Su Tutma Yüzde Değerleri için Parametre Tahmin Değerleri.

Parametre Tahmin Değerleri			
Parametre	Tahmini	St. Sapma	t-değer
AR(1)	0.6032	0.0970	6.22
Sabit	23.0379	0.2250	102.37
Ortalama	58.0584	2.501	
HKT	254.969		
HKO	3.642		
Box-Pierce	24 Lag	32.80	
Tablo Değ.	23 s.d	35.17	
Tablo değ > Box-Pierce old. model uygun.			

Tablo-4.42, Gıda Sektöründe Unların Su Tutma Yüzde Değerleri için Sapan Değer Analizi Sonuçları.

Sapan Değer Analizi Sonuçları-1		
Parametreler	Sapan Değer	Döngü
AR 0.6032	AO 23.Pozisyon	1

Üçüncü tip verilere uygulanan sapan değer analizi sonucunda 23'üncü pozisyonda birinci tip sapan değer görülmüştür. Veriler incelenecek olursa; 22'inci gözlem değeri 56.28 iken, 23'üncü gözlem değerinin 63.17 değerine çıktığı daha sonra 24'üncü gözlem değeri 58.79 değerini almıştır. Bir önceki gözlem değerine bağlı olarak hesaplanan 23'üncü gözlem değeri 22'inci gözlem değerinden büyük olduğundan sapan olarak elde edilmiştir. 23'üncü pozisyonda bulunan 63.17 değerine karşılık 56.58246 değeri hesaplanarak veri değiştirilmiş, yeni verilere sapan değer analizi uygulanmıştır.

Tablo-4.43, Gıda Sektöründe Unların Su Tutma Yüzde Değerleri için Sapan Değer Analizi Sonuçları.

Sapan Değer Analizi Sonuçları-2		
Sapan Değer Etkileri Giderildikten Sonra		
Parametreler	Sapan Değer	Döngü
AR 0.6625	AO Rastlanmadı	2

Sapan değer etkisi giderildikten sonra uygulanan analiz işlemi sonucunda sapan değere rastlanmamıştır.

BÖLÜM 5

5. SONUÇLAR

Yapılan çalışma ile elde edilen sonuçlar paragraflar halinde aşağıda ifade edilmiştir :

Zaman serilerinde ortaya çıkması muhtemel sapan değerleri tespit amacıyla kullanılan tarama süreçleri, otokorelasyona dayalı süreçlerdir. Bu nedenle, doğal rasgelelik sonucunda ortaya çıkan, bir başka deyişle seride otokorelasyon düzensizliğine sebep olmayan ikinci tip sapan değer (IO) tespit edilme olasılığı; ölçüm, tartım, kaydetme ve benzeri sebeplerden meydana gelen birinci tip sapan değere (AO) oranla daha yüksek bulunmaktadır. Sonuçları görmek için simülasyon tablolarına veya bu tabloların kullanımı ile elde edilen varyans analizi sonuçlarına bakılabilir.

Zaman serisi modellerinin (AR, MA), sapan değer tespitindeki etkileri için yapılan analiz işlemi sonucunda modellerin tespitinde etkili olduğu sonucunu ortaya koymuştur. Simülasyon tablolarından ve varyans analizi sonuçlarından edinilen bilgiye göre, MA modeli ile ifade edilen gözlem setleri için sapan değer tespit olasılığı daha yüksektir. Bilindiği üzere MA modellerde gözlem değerleri, kendinden önceki gözlem değerlerine ilişkin hata terimlerinin doğrusal kombinasyonu şeklinde ifade edilmektedir. Bunun sonucu olarak seriyi oluşturan gözlem değerleri arasında otokorelasyon daha kuvvetli olmakta ve seride sapan değer bulunması durumunda otokorelasyona dayalı tarama süreçleri tarafından daha kolay bir şekilde belirlenmektedir.

Sapan değer türlerinin (AO, IO) sapan değer tespitinde etkileri üzerine yapılan analiz işlemi sonucunda, sonuç önemli bulunmuştur. Bu durumda ikinci tip sapan değer (IO) tarama süreçleri tarafından daha yüksek bir olasılıkla tespit edildiği sonucu elde edilmiştir. Birinci tip sapan değer (AO) durumunda, gözlemin kendisi etkilenmekte, kendinden sonraki gözlemler etkilenmemektedir. Böylece kendinden önce ve sonra gelen gözlemler arasında oluşan korelasyon kaybı, gözlemin tarama süreci tarafından tespitini güçleştirmektedir.

Sapan değeri büyüklüğüne (3σ , 5σ) ilişkin analiz işlemi sonucunda sapan değeri büyüklüğünün tespitinde önemli olduğunu ortaya koymuştur.

Seri uzunluklarının (50, 100, 150) sapan değeri tespitinde etkisini ortaya koymak için yapılan analiz işlemi sonucunda 50 ile 100 gözlem değeri arasında fark önemli bulunurken, 100 ile 150 seri uzunlukları için sapan değeri tespitinde farklılık görülmemektedir. Bu durumda seri uzunluğunun 100 ve daha büyük olduğu durumlar için sapan değeri tespitinde bir farklılık meydana gelmemektedir.

Sapan değeri tespitinde ölçüt değerlerinin etkisi de irdelenmiştir. Yüksek duyarlı testler için $C=3.00$, orta duyarlı testler için $C=3.50$, düşük duyarlı testler için kullanılan $C=4.00$ ölçüt değerlerinin sapan değeri tespitinde istatistiksel bakımdan farklılık yaratmadığı, dolayısıyla duyarlılığın sapan değeri tespitinde önemsiz bulunduğu sonucuna varılmıştır.

Varyans analizi sonuçları incelendiğinde; sapan değeri büyüklüğü, sapan değeri türü ve zaman serisi modellerinin sapan değeri tespitinde etkili olduğu sonucu gözlemlenebilir. Seri uzunluğunun da 50 ile 100 arası önemli 100'den büyük serilerin analizi için bir farklılığın söz konusu olmadığı görülmüştür. Yapılan Çıkarımlar, Ayrıntılı tablolardan görülebilir. Aynı analiz işlemi sonucunda Model-Büyüklik, Model-Tür, Büyüklik-Tür, Genişlik-Tür ve Duyarlılık-Tür ikili etkileşimleri önemli bulunmuştur. Model, Büyüklik, Genişlik, Duyarlılık ve Tür unsurlarından oluşan iki'li ve üç'lü etkileşimlerden modele dahil edilenler önemsiz bulunmuştur. Tek sapan değeri için elde edilen sonuçlar, çift sapan değeri için de aynen elde edilmiştir. Ancak belirtilmesi gerekli önemli bir husus, gözlem seti içerisinde sapan değeri sayısının birden ikiye çıkması tarama sürecinin gücünde büyük oranda zayıflamaya sebep olmuştur.

Sapan değeri tarama süreçlerinin sapan değeri tespitinde etkilerini tespit etmek amacıyla yapılan varyans analizi işlemi sonucunda, tarama süreçlerinin tespitinde etkili olduğu sonucuna varılmıştır. Böylece LS, M-H, M-B, GM-H, GM-B ve ARDIŞIK yöntemleri içerisinde Hata Kareler Ortalaması minimum elde edilen ARDIŞIK yöntemin

en etkin yöntem olduđu sonucuna varılmıřtır. Aynı model denklemi üzerinde incelenen sapan deęer turleri, ve sapan deęer sayısı da bryk bir oranda onemli bulunmuřtur. Turl, Surlreç ve Adet (sapan deęer sayısı) fakturlerinden oluřan iki'li etkileřimlerin onemli bulunduđu sonucuna ulařılmıřtır.

Kalite kontrol amaçlı sapan deęer analizi çalıřmalarında, uretim surlrecinde elde edilen gurlzlemlere iliřkin olçum tartım ve benzeri ozelliklerin homojen yapı gurlstermesi gerekir. Gurlzlem deęerlerinin homojen olmaması durumunda hata kaynaęı saęlıklı bir řekilde tespit edilemeyebilir. Bu nedenle sapan deęer analizi ilaç uretimi gibi pahalı uretim surlreçleri için ekonomik bir kontrol mekanizması olmayabilir. Bunun yerine uretim anında kalite kontrol iřlemini yapan taguchi yontemleri gibi etkin yontemlere bařvurmak daha etkili olabilir.

Eldeki ornoklem deęerleri için bir model denklemi onęoruluyorsa, uygun model denkleminin seçilmesi son derece onemlidir. Eęer model denklemi uygun deęilse, sapan olmayan gurlzlem deęerleri sapan olarak elde edilir. Bu da analizin yanlı sonuçlar doęurmasına sebep olabilir.

Bir veri setinde gurlzlem deęeri; olçum, tartım, kaydetme ve benzeri gibi kiři veya aletlerden kaynaklanmış ise, sapan olan gurlzlemin seriden silinmesi veya uygun bir test yontemi ile durlzeltmesi gereklidir. Gurlzlem deęeri, doęal rasgelelik sonucunda sapan olmuş ise, gurlzlem deęerini seriden çıkarmak veya durlzeltme yontemine tabi tutarak veriyi durlzeltmek kiřinin tercihine bırakılmalıdır.

Birinci uygulamada, Box-Jenkins A serisi olarak bilinen ve kalite kontrol amaçlı surlreçlerin analizine yonelik kullanılan, iki saatte bir elde edilmiş kimyasal konsantrasyon olçumlerinden oluřan veriler kullanılmıştır. Yapılan analiz iřlemi sonucunda, verilerde biri birinci tip, dięeri ikici tip olmak uzere iki adet sapan deęere rastlanmış, sapan deęerlere iliřkin yorumlar yapıldıktan sonra, sapan etkiler giderilmiş, yapılan ikinci analiz iřlemi sonucunda sapan deęere rastlanmamıřtır.

İkinci uygulamada düzeltme amaçlı İzmir ÇİMENTAŞ T.A.Ş' nin 1980-1987 yılları arası yapılan çimento ihracat verileri kullanılmıştır. Söz konusu veriler üzerinde yapılan analiz işlemi sonucunda tek sapan değere rastlanmıştır, sapan değer etkisi giderildikten sonra yapılan ikinci analiz işlemi sonucunda tahmine ilişkin hata kareler ortalaması 189'dan 167'ye gerileyerek % 12 küçülmüştür. Bu sonuç elde edilen yeni parametre değerinin etkinliği yönünden önemlidir.

Etkinlik artırma amaçlı yapılan üçüncü uygulamada sapan değer analiz işlemi sonucunda zaman serileri analiz parametreleri büyük oranda değişim göstermiştir. Sapan etki giderildikten sonra 0.6032 olarak elde edilen AR parametre değeri, 0.6625 olarak, parametreye ilişkin elde edilen standart sapma değeri 0.0970 iken 0.0916 olarak % 10'luk bir küçülme göstermiş, parametreye ilişkin elde edilen t-hesap değeri 6.22 iken 7.23 olarak değişmiştir. Modele ilişkin Box-Pierce ki-kare değeri 32.80 iken 24.60 olmuş ve seri için uydurulan model denklemini, daha gerçeğe yakın bir özellik kazanmıştır. Benzer şekilde hata kareler toplamı 254.969 iken bu değer, 212.657'ye gerilemiştir. Hangi yönden ele alınıralsa alınsın, kişilerden, ölçüm, tartım ve kaydetme hatalarından meydana gelen birinci tip sapan değerlerin etkilerini gidermek gerekmektedir.

Sapan değerler kavramı zaman serilerinde inceleniyorsa, gizlenme ve sürüklenme etkilerinin kendini göstermesi nispeten zayıftır. Çünkü, zaman serisi gözlem değerleri, bir gözlemin kendinden önceki gözlem değerlerinin, kendinden önceki hata değerlerinin veya kendinden önceki gözlem ve hata terimlerinin doğrusal kombinasyonu olarak elde edilirler. Buna göre serileri oluşturan gözlem değerleri arasında bağımlılık meydana gelir. Gözlem değerleri arasındaki bağımlılık hata terimleri üzerinde de bağımlılığı, dolayısıyla otokorelasyon sonucunu doğurmaktadır. Hata terimleri arasındaki otokorelasyonun kuvvetli olması gizlenme ve sürüklenme nedeniyle sapan değer oluşumunu engelleyen bir unsur gibi görünmektedir.

BİBLOGRAFYA VE EKLER

- Biblografya.
- EK-1, Ardışık Yöntem için Bilgisayar Programı.
- EK-2, Dağılımları Verilen Örneklemeler için Sapan Değer İstatistikleri.



BİBLOGRAFYA

1. **Abraham, B ve Chuang, A.**, 1989: Outlier detection and time series modelling, *Technometrics*, 31,241-248.
2. **Abraham, B ve Yatawara, N.**, 1988: A score test for detection of time series outliers. *J. Time Ser. Anal.*,9, 109-119.
3. **Abraham, B. ve Box, G. E. P.**, 1979: Bayesian analysis of some outlier problems in time series, *Biometrika*, 66 , 2 , p : 229-236, Printed in Great Britain.
4. **Acar, W. ve Booth, P. D. E.**, 1987: Easier Quality Control: Combaining problem classification with time series analysis: Graduate School of Managment, Kent state University, Kent, oh 44242-0001.
5. **Akpınar, H.**, 1995: Excel' de fonksiyonlar, veri Analizleri ve problem çözmeye, İstanbul Üniversitesi, İşletme Fakültesi Yayın No. 249.
6. **Al-Khalidi, A. S. R. S.**, 1997: Using intervention analysis in time series to measure water sources pollution, III. Ulusal Ekonometri ve İstatistik Sempozyumu, Bursa.
7. **Alba, E ve Zartman, L. D.**. 1980: Testing outliers in time series, *Anal. Time Ser*, North-Holland publishing Company.
8. **Andrews, D. F. ve Pregibon, D.**, 1978: Finding the outliers that matter, *J.R. Stat. Soc. Ser B*, 40: 85-93.
9. **Anderson, D. O.**, 1980: Testing Outliers in Time Series: An application to remotely sensed tempratures in cattle. *Analysing Time Series*, North-Holland Publishing Company.
10. **Anscombe M. ve Tukey, J. W.**, 1963: The examination and analysis of residuals, *Technometrics*, 5, 141.

33. **Muirhead, C. R.**, 1986: Distinguishing outlier types in time series., J.R. Statist. Soc. B. 48, NO:1 S:39-47.
34. **Muluk, Z., Kurt, S., Toktamış, Ö., Karaağaoğlu, E.**, 1994: Deney Düzenlemede İstatistik Yöntemler, Ege Üniversitesi Fen Fakültesi Yayınları No. 146, Ege Üniversitesi Basımevi, Bornova-İzmir.
35. **Özmen, A.**, 1986: Zaman serisi analizinde Box-Jenkins yöntemi ve Banka mevduat tahmininde uygulama denemesi., Fen Edebiyat Fakültesi yayınları No: 9., Eskişehir.
36. **Pena, D.**, 1987: Measuring the importance of outliers in ARIMA models in New perspectives in theoretical and applied statistics., New york: Wiley.
37. **Pena, D. ve Guttman, I.**, 1993: Comparing Probabilistic Methods for Outlier Detection in linear Models, Biometrika, 80, 3, 603-10.
38. **Prins, J.**, 1980: Interactive multivariate time series analysis., Anal. Time Ser., North-holland Publishing Company.
39. **Pukkila, T., Koreisha, S. ve Kallinen, A.**, 1990: The identification of ARMA models., Biometrika, 77, s:537-548., Printed in Great Britain.
40. **Senta, E. ve Kounias, S.**, 1990: Parameter estimation and order determination of otoregressive models, Anal. Time Ser, North-Holland Publishing Company.
41. **Sharples, D.**, 1990: Identification and accomodation of outliers in general hierarchical models., Biometrika, 77, s:445-453.
42. **Shiao, L. Wu, Y. ve Hosking, J. R. M.**, (1993) Reallocation outliers in time series., Applied Statistics., 42, No:2. S:301-313.

- 43.Şen, A. ve Şehirliođlu, A. K., 1993: Sapan ve etkili gözlemlerin bulunmasında kullanılan istatistikler, Ekonometri- İstatistik Sempozyumu, İzmir.
- 44.Şenyay, L., 1986: Zaman serisi analizlerinin temel prensipleri ve bu analiz için Genstat paket programının kullanımı üzerine bir çalışma., Ege Üniversitesi Bilgisayar Araştırma ve Uygulama Merkezi (BAUM) Dergisi., Cilt : 9, Sayı : 2.
- 45.Taplin, H. R., 1993: Robust likelihood calculation for time series., J.R Statist. Soc., b, 55, No:4 s:829-836.
- 46.Tatlıdil, H., 1981: Doğrusal regresyonda ve çok değişkenli verilerde kuşkulu gözlemlerin testi., Basılmamış doktora tezi., Hacettepe Üniversitesi., Ankara.
- 47.Tecim, V., 1990: Zaman serisi analizlerinde sermaye piyasası uygulamaları, Basılmamış yüksek lisans tezi., D. Eylül Üniv., İzmir.
- 48.Tsay, R.S., 1988: Outliers, level shifts, and variance changes in time series : Journal of forecasting., 7: 1-20.
- 49.Wheelright, S.C ve Makridakis, S., 1973: Forecasting methods for managment., John Wiley & Sons Inc., New york, 1973.
- 50.William R. Bell., 1994 : U.S Bureau of the Cencus, A compute program for detecting outliers in time series, The British Council Library, Gaziosmapaşa-Ankara.

EK-1.

ARDIŞIK YÖNTEM İÇİN BİLGİSAYAR PROGRAMI

```
10 REM BU PROGRAM ZAMAN SERİLERİNDE SAPAN DEĞER ANALİZİ YAPAR
20 REM Bu Program Zaman Serilerinde Ortaya Çıkması Muhtemel
30 REM Sapan Değerlerin Pozisyon ve Tiplerini Ortaya Çıkarr.
40 REM Yüksek Duyarlılık İçin c=3.00, Orta Duyarlılık İçin c=3.50 ve
50 REM Düşük Duyarlılık İçin c=4.00 alınmalıdır.
60 DIM HATA(197), PI(197), FMBDA(2,197), Z(197)
70 DIM X(197), ZO(5), ZS(5), WA(197), WI(197)
80 LET L=0: KEY OFF:CLS
90 REM Model Parametrelerinin Okutulması
100 PRINT "ARIMA(p,d,q) Parametrelerini Okutunuz"
110 PRINT "p (AR) Parametresini Giriniz "
120 INPUT "", P
130 PRINT"d (FARK) Parametresini Giriniz"
140 INPUT "",D
150 PRINT "q (MA) parametresini Giriniz"
160 INPUT "", Q
170 PRINT "Gözlem Sayısını Giriniz"
180 INPUT "", N
190 PRINT "Gözlemleri Giriniz"
200 AS=INPUT$(1)
210 FOR K=1 TO N
220 LOCATE 10,10:PRINT K," Gözlem"
230 LOCATE 10,25:INPUT "",Z(K)
240 CLS
250 NEXT
260 REM Hata Değerlerinin Okutulması
270 FOR K=1 TO N
280 LOCATE 10,10:PRINT K,".Hata"
290 LOCATE 10,25:INPUT "", HATA(K)
300 NEXT
310 REM Pi Ağırlıklarının Okutulması
320 FOR K=1 TO N
330 LOCATE 10,10:PRINT K,".pi Değeri"
340 LOCATE 10,25:INPUT "",PI(K)
350 NEXT
360 LET I=197
370 REM Serinin Tahmin Edilen Parametre Değerleri
380 REM Hata varyansı ve c sabitinin okutulması
390 PRINT "Tahmini Parametre Değerlerini Giriniz"
400 PRINT "AR Parametre Değerlerini Giriniz"
410 INPUT INPUT "",FI
420 PRINT "MA Parametre Değerini Giriniz"
430 INPUT "", TETA
440 PRINT "Varyans Değerini Giriniz"
```

```

450 INPUT "",VAR
460 PRINT"C Parametre Deęerini Giriniz"
470 INPUT "",C
480 STHATA=VAR^0.5
490 FOR K2=1 TO 2
500 FOR K3=1 TO N
510 FMBDA(K2,K3)=0
520 NEXT
530 NEXT
540 FOR K=1 TO N
550 FMBDA(1,K)=HATA(K)/STHATA
560 IF K=1 THEN 590
570 IF ABS(FMBDA(1,K))>ENB1 THEN 590
580 GOTO 620
590 ENB1=ABS(FMBDA(1,K))
600 T1=1
610 T2=K
620 NEXT
630 REM Omegaa' nın Bulunması
640 FOR K=1 TO N-1
650 T4=N-K
660 TP1=HATA(K)
670 TP2=1
680 FOR K2=1 TO T4
690 F1=PI(K2)*HATA(K+K2)
700 FM=PI(K2)^2
710 TP1=TP1-F1
720 TP2=TP2+FM
730 NEXT
740 OMEGAA=TP1/TP2
750 REM Lambda(2,T)'lerin ve en büyük lambda(2,t)'nin bulunması ,
760 FMBDA(2,K)=(OMEGA*TP2^0.5)/STHATA
770 IF K=1 THEN 800
780 IF ABS(FMBDA(2,K))>ENB2 THEN 800
790 GOTO 830
800 ENB2=ABS(FMBDA(2,K))
810 T3=2
820 T4=K
830 NEXT
840 REM Lambda(1,t) ve Lambda(2,t)'ler arasından c'den büyük olanın bulunması
850 IF (ENB1>ENB2) AND ENB1>C THEN 1080
860 IF (ENB2>ENB1) AND ENB2>C THEN 890
870 PRINT "Sapan Deęere Rastlanmadı !..."
880 GOTO 1240
890 REM 1. Tip sapan deęer etkisinin bulunması
900 TP1=HATA(T4)
910 TP2=1

```



```

920 FOR K=1 TO N-T4
930 F1=PI(K)*HATA(T4+K)
940 FM=PI(K)^2
950 TP1=TP1-F1
960 TP2=TP2-FM
970 NEXT
980 OMEGAA=TP1/TP2
990 LET L=L+1
1000 ZS(L)=T4
1010 ZO(L)=1
1020 WA(T4)=OMEGAA
1030 PRINT "1. Tip sapan değere rastlandı !..."
1040 HATA(K)=HATA(K)-OMEGAA*(-PI(K-T4))
1050 NEXT
1060 GOTO
1070 REM 2. Tip sapan değeri var. Hatalar yeniden hesaplanıyor .
1080 OMEGAI=HATA(T2)
1090 HATA(T2)=0
1100 L=L+1
1110 ZS(L)=T2
1120 ZO(L)=2
1130 WI(T2)=OMEGAI
1140 PRINT "2. Tip sapan değere rastlandı !..."
1150 PRINT OMEGAI, ENBI
1160 TP=0
1170 FOR K=1 TO N
1180 THATA=HATA(K)^2
1190 TP=TP+THATA
1200 NEXT
1210 REM Yeni hata varyansının hesaplanması
1220 VAR=TP/N
1230 GOTO 480
1240 IF L=0 THEN 1360
1250 FOR I=1 TO L
1260 M=ZS(I)
1270 IF ZO(I)=1 THEN X(M)=Z(M)-WA(M):GOTO 1320
1280 IF P=0 AND D=1 AND Q=1 THEN X(M)=Z(M)-Z(M-1)+X(M-1)-
WI(M)+TETA*WI(M-1)
1290 IF P=1 AND D=1 AND Q=1 THEN ZS(I)=M:X(M)=Z(M)-FI*Z(M-1)+FI*X(M-
1)+TETA*WI(M-1)+FI*X(M-1)-FI*X(M-2)
1300 IF P=1 AND D=0 AND Q=1 THEN X(M)=Z(M)-FI*Z(M-1)+FI*X(M-1)+TETA*WI(M-
1)-WI(M)
1310 IF P=1 AND D=0 AND Q=0 THEN X(M)=Z(M)-WI(M)+FI*WI(M-1) ELSE
X(M)=Z(M)-WI(M)+TETA*WI(M-1)
1320 PRINT M, X(M)
1330 FOR K=1 TO N
1340 PRINT X(K)
1350 NEXT
1360 END

```

EK-2.

DAĞILIMLARI VERİLEN ÖRNEKLER İÇİN SAPAN DEĞER İSTATİSTİKLERİ

Tablolara ilişkin simge ve gösterimler :

N μ ve σ^2 her ikisi de bilinmiyor.

N ν μ bilinmiyor, σ^2 ise νS_v^2 , χ_v^2 dağılımlı ve örneklemden bağımsız iken $\nu = S_v^2$ gibi bir tahmin değeridir.

N μ μ biliniyor, σ^2 bilinmiyor.

N σ σ^2 biliniyor, μ bilinmiyor.

N $\mu\sigma$ μ ve σ^2 'nin her ikisi de biliniyor.

G Merkezi bilinmeyen gama dağılımlı örneklem.

E Merkezi bilinmeyen üstel dağılımlı örneklem.

G a Merkezi "0" da (ya da genel olarak a' da) olan gama dağılımlı örneklem.

E a Merkezi "0" da (ya da genel olarak a' da) olan üstel dağılımlı örneklem.

<u>Dağılımları Verilen Örneklerdeki</u>		
<u>Kuşkulu Gözlem Değerlerinin Testinde Kullanılan İstatistikler</u>		
Simge	Testin Açıklaması	Test İstatistiği
N1	$x_{(n)}$ üst değerinin testi	$\frac{x_{(n)} - \bar{x}}{S}$ ya da $\frac{S_n^2}{S^2}$
N2	Uçlardaki değer testi	$\max\left(\frac{x_n - \bar{x}}{S}, \frac{\bar{x} - x_{(1)}}{S}\right)$ ya da $\min\left(\frac{S_n^2}{S^2}, \frac{S_1^2}{S^2}\right)$
N3	$x_{(n-k+1)}, \dots, x_{(n)}$ ($k > 2$) gibi üst değer test	$(x_{(n-k+1)} + \dots + x_{(n)} - k\bar{x}) / S$
N4	$x_{(n-k+1)}, \dots, x_{(n)}$ ($k > 2$) gibi üst değer test	$(S_{n-k+1}^2, \dots, S_{n-1}^2) / S^2$
N5	$x_{(1)}, x_{(n)}$ alt ve üst değer çifti için test	$S_{1,n}^2 / S^2$
N6	$x_{(1)}, x_{(n)}$ alt ve üst değer çifti için test	$(x_{(n)} - x_{(1)}) / S$
N7 (Nμ7)	$x_{(n)}$ üst değeri için Dixon türü test	$\frac{x_{(n)} - x_{(n-1)}}{x_{(n)} - x_{(1)}}$
N8 (Nμ8)	Uçlardaki değerler için Dixon türü test	$\max\left(\frac{x_{(n)} - x_{(n-1)}}{x_{(n)} - x_{(1)}}, \frac{x_{(2)} - x_{(1)}}{x_{(n)} - x_{(1)}}\right)$
N9 (Nμ9)	$x_{(n)}$ üst değeri için Dixon türü test	$\frac{x_{(n)} - x_{(n-1)}}{x_{(n)} - x_{(2)}}$
N10(Nμ10)	$x_{(n)}$ üst değeri için Dixon türü test	$\frac{x_{(n)} - x_{(n-1)}}{x_{(n)} - x_{(3)}}$
N11(Nμ11)	$x_{(n-1)}, x_{(n)}$ üst değer çifti için Dixon türü test	$\frac{x_{(n)} - x_{(n-2)}}{x_{(n)} - x_{(1)}}$
N12(Nμ12)	$x_{(n-1)}, x_{(n)}$ üst değer çifti için Dixon türü test	$\frac{x_{(n)} - x_{(n-2)}}{x_{(n)} - x_{(2)}}$

Simge	Testin Açıklaması	Test İstatistiği
N13(N μ 13)	$x_{(n-1)}, x_{(n)}$ üst değer çifti için Dixon türü test	$\frac{x_{(n)} - x_{(n-2)}}{x_{(n)} - x_{(3)}}$
N14	Bir ya da daha çok üst değer için test	Örneklem Çarpıklığı $g_1 = \sqrt{b_1}$
N15	Yönlerine bakılmaksızın bir yada daha çok değer için iki yönlü test	Örneklem Basıklığı b_2
N16	Yönlerine bakılmaksızın k değer için blok testi	Grubbs-Türü İstatistik E_k
N17	Belirsiz sayıda değer olması durumunda	Shapiro ve Wilks' in W-İstatistiği
Nv1	$x_{(n)}$ üst değerinin testi	$\frac{x_{(n)} - \bar{x}}{S_v}$
Nv2	$x_{(n)}$ üst değerinin testi	$\frac{x_{(n)} - \bar{x}}{S}$
Nv3	Uçlardaki değerlerin testi	$\max\left(\frac{x_{(n)} - \bar{x}}{S_v}, \frac{\bar{x} - x_{(1)}}{S_v}\right)$
Nv4	Uçlardaki değerlerin testi (Nv2' nin iki yönlü biçimi)	$\max\left(\frac{x_{(n)} - \bar{x}}{S}, \frac{\bar{x} - x_{(1)}}{S}\right)$
Nv5	$x_{(n-k+1)}, \dots, x_{(n)}$ gibi ($k > 2$) üst değer testi	$\frac{x_{(n-k+1)} + \dots + x_{(n)} - k\bar{x}}{S}$
Nv6	$x_{(1)}, x_{(n)}$ alt ve üst değer çiftinin testi	$\frac{x_{(n)} - x_{(1)}}{S_v}$
Nv7	$x_{(1)}, x_{(n)}$ alt ve üst değer çiftinin testi	$\frac{x_{(n)} - x_{(1)}}{S}$
N μ 1	$x_{(n)}$ üst değerinin testi (N μ 1'in iki yönlü biçimi)	$\frac{x_{(n)} - \mu}{S(\mu)}$

Simge	Testin Açıklaması	Test İstatistiği
N μ 3	$x_{(n-k+1)}, \dots, x_{(n)}$ gibi ($k > 2$) üst değer testi	$\frac{x_{(n-k+1)} + \dots + x_{(n)} - k\mu}{S(\mu)}$
N μ 4	$x_{(n-1)}, x_{(n)}$ üst değer çiftinin testi	$S_{n-1}^2(\mu) / S^2(\mu)$
N μ 5	$x_{(1)}, x_{(n)}$ alt ve üst değer çiftinin testi	$S_{1,n}^2(\mu) / S^2(\mu)$
N μ 6	Bir ya da daha çok uç değer için iki yönlü test	$\frac{\sum_{j=1}^n (x_j - \mu)^4}{nS^4(\mu)}$
N μ 14	Yönelim bakılmaksızın k değer için blok test	$\frac{\sum_{j=n-k+1}^n x_j - \mu }{S^2(\mu)}$
N σ 1	$x_{(n)}$ üst değerinin testi	$\frac{x_{(n)} - \bar{x}}{\sigma}$
N $\mu\sigma$ 1	$x_{(n)}$ üst değerinin testi	$\frac{x_{(n)} - \mu}{\sigma}$
N σ 2	Uçlardaki değerlerin testi (N σ 1 'in iki yönlü biçimi)	$\max\left(\frac{x_{(n)} - \bar{x}}{\sigma}, \frac{\bar{x} - x_{(1)}}{\sigma}\right)$
N σ 3	$x_{(n-k+1)}, \dots, x_{(n)}$ gibi ($k > 2$) üst değer testi	$\frac{x_{(n-k+1)} + \dots + x_{(n)} - k\bar{x}}{\sigma}$
N $\mu\sigma$ 3	$x_{(n-k+1)}, \dots, x_{(n)}$ gibi ($k > 2$) üst değer testi	$\frac{x_{(n-k+1)} + \dots + x_{(n)} - k\mu}{\sigma}$
N σ 4	$x_{(n-1)}, x_{(n)}$ üst değer çiftinin testi	$S_{n-1,n}^2 / \sigma^2$
N $\mu\sigma$ 4	$x_{(n-1)}, x_{(n)}$ üst değer çiftinin testi	$S_{n-1,n}^2(\mu) / \sigma^2$
N σ 5	$x_{(1)}, x_{(n)}$ alt ve üst değer çiftinin testi	$S_{1,n}^2 / \sigma^2$
N $\mu\sigma$ 5	$x_{(1)}, x_{(n)}$ alt ve üst değer çiftinin testi	$S_{1,n}^2(\mu) / \sigma^2$

Simge	Testin Açıklaması	Test İstatistiği
Nσ6(Nμσ6)	$x_{(1)}, x_{(n)}$ alt ve üst değer çiftinin testi	$\frac{x_{(n)} - x_{(1)}}{\sigma}$
Nσ7(Nμσ7)	$x_{(n)}$ üst değerinin testi	$\frac{x_{(n)} - x_{(n-1)}}{\sigma}$
Nσ8(Nμσ8)	$x_{(n-1)}, x_{(n)}$ üst değer çiftinin testi	$\frac{x_{(n-1)} - x_{(n-2)}}{\sigma}$
Nσ9(Nμσ9)	k alt ve k üst değer testi	$\frac{x_{(n-k+1)} - x_{(k)}}{\sigma}$
Ga1(Ea1)	Gamma (üstel) dağılımlı bir örnekleme $x_{(n)}$ üst değerinin testi	$\frac{x_{(n)}}{\sum x_j}$
Ea2	Üstel dağılımlı bir örnekleme $x_{(n)}$ üst değerinin testi	$\frac{x_{(n)} - x_{(n-1)}}{x_{(n)}}$
E2	Üstel dağılımlı bir örnekleme merkezine bakılmaksızın $x_{(n)}$ üst değerinin testi	$\frac{x_{(n)} - x_{(n-1)}}{x_{(n)} - x_{(1)}}$
Ga3(Ea3)	Gamma (üstel) dağılımlı bir örnekleme $x_{(1)}$ alt değerinin testi	$\frac{x_{(1)}}{\sum x_j}$
E4	Merkezi bilinmeyen üstel dağılımlı bir örnekleme $x_{(1)}$ alt değerinin testi	$\frac{x_{(2)} - x_{(1)}}{x_{(n)} - x_{(1)}}$
Ga5(Ea5)	Gamma (üstel) dağılımlı bir örnekleme $x_{(n-k+1)}, \dots, x_{(n)}$ gibi ($k > 2$) üst değer testi	$\frac{x_{(n)} + \dots + x_{(n-k+1)}}{\sum x_j}$
Ea6	Üstel dağılımlı bir örnekleme $x_{(n)}, x_{(n-1)}$ üst değer çiftinin testi	$\frac{x_{(n)} - x_{(n-1)}}{x_{(n)}}$
E6	Üstel dağılımlı bir örnekleme merkeze bakılmaksızın $x_{(n)}, x_{(n-1)}$ üst değer çiftinin testi	$\frac{x_{(n)} - x_{(n-2)}}{x_{(n)} - x_{(1)}}$

Simge	Testin Açıklaması	Test İstatistiği
Ga7(Ea7)	Gamma (üstel) dağılımlı bir örnekleme $x_{(1)}, x_{(n)}$ alt ve üst değer çiftinin testi	$\frac{x_{(n)}}{x_{(1)}}$
E8	Merkezi bilinmeyen üstel dağılımlı bir örnekleme $x_{(1)}, x_{(n)}$ alt ve üst değer çiftinin testi	$\frac{x_{(n-1)} - x_{(2)}}{x_{(n)} - x_{(1)}}$
Ga9(Ea9)	Gamma (üstel) dağılımlı bir örnekleme $x_{(1)}, \dots, x_{(k)}$ gibi ($k > 2$) alt değerlerin testi	$\frac{x_{(1)}, \dots, x_{(k)}}{\sum x_j}$
E10	Merkezi bilinmeyen üstel dağılımlı bir örnekleme $x_{(1)}, x_{(2)}$ alt değer çiftinin testi	$\frac{x_{(3)} - x_{(1)}}{x_{(n)} - x_{(1)}}$
Ea11	Üstel dağılımlı bir örnekleme merkez bilgisinin kullanıldığı Dixon türü test	$\frac{x_{(s)} - x_{(r)}}{x_{(q)} - a}$
E11	Üstel dağılımlı bir örnekleme a merkez bilgisinin kullanılan Dixon türü test	$\frac{x_{(s)} - x_{(r)}}{x_{(q)} - x_{(p)}}$
E12	Üstel dağılımlı bir örnekleme belirlenmemiş sayıda değerlerin testi	$\frac{n(\bar{x} - x_{(1)})}{(n-1) \sum_{j=1}^n (x_j - \bar{x})^2}$
Ga13	Gamma dağılımlı bir örnekleme r şekil parametresi için test	N1-N17